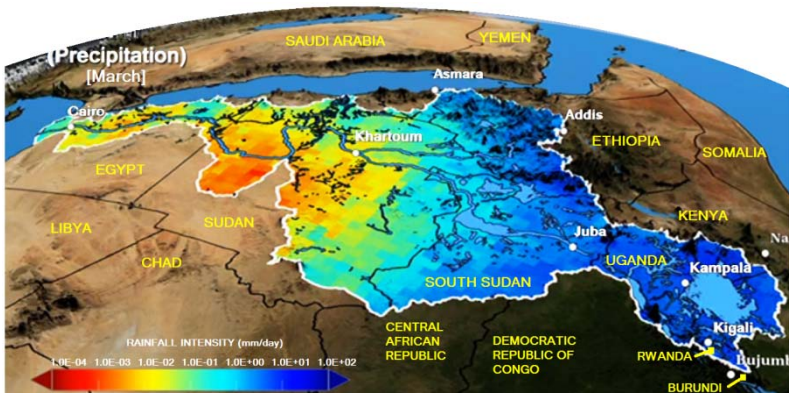


Variability of Rainfall and River Flow in the Nile Basin



Charles Onyutha

Dissertation presented in partial fulfilment of
the requirements for the degree of Doctor
of Engineering Science (PhD):
Civil Engineering

Supervisor:

Prof. Patrick Willems

September 2017

Variability of Rainfall and River Flow in the Nile Basin

Charles ONYUTHA

Supervisor:

Prof. Patrick Willems

Members of Examination Committee:

Prof. Hilde Heynen (Chairperson)

Prof. Jaak Monbaliu

Prof. Jozef Van Dyck

Prof. Charles Niwagaba (Makerere University, Uganda)

Ass. Prof. Wim Thiery (Vrije Universiteit Brussel)

Dissertation presented in partial
fulfilment of the requirements for
the degree of Doctor of Engineering
Science (PhD): Civil Engineering

September 2017

The cover image showing the long-term averaged precipitation of March, originally had no labels of the visible countries, and was adopted from a video of the Goddard Media Studios on "Distributed Water Balance of the Nile Basin" available online via the link <http://svs.gsfc.nasa.gov/11223> (accessed 22.08. 2017).

*Psalm 147:8
Our God covers the heavens with clouds,
giving rain for the earth and making grass grow on the hills.*

*Job 5:10
He gives rain on the earth and sends waters on the fields.*

*James 5:18
Then he prayed again, and the sky gave rain,
and the earth bore its fruit.*

*"Predicting rain doesn't count.
Building arks does"
Warren Buffett*

*Matthew 24:38-39
"For as in those days
before the flood they were eating
and drinking, marrying and giving in marriage,
until the day that Noah entered the ark, and they did not understand
until the flood came and took them all away; so will the coming of the Son of Man be"*

© 2017 KU Leuven – Arenberg Doctoral School of Science, Engineering & Technology,
Kasteelpark Arenberg 40 box 2448, B-3001 Leuven

Alle rechten voorbehouden. Niets uit deze uitgave mag, worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaandelijke schriftelijke toestemming van de uitgever.

All rights are reserved. No part of this publication may be reproduced in any form by print, photo print, microfilm, electronic or any other means without the written permission of the publisher.

Dedication

I dedicate this work to

Paula (my wife) and

Lina, Lincoln, and Eli Lelisa (my children).

Acknowledgements

It is so gratifying when the wanderings in a challenging journey are actually the unfolding of a secret preordained destination. This is why I am grateful to God's unconditional faithfulness, grace and providence in empowering me to successfully complete this research. I am thankful to Prof. Patrick Willems for his academic mentorship and supervision. I acknowledge the financial support of the IRO PhD scholarship of KU Leuven.

My heart is besieged with respect and gratitude for my caring wife Paula and children Lina, Lincoln, and Eli Lelisa for their support, and above all the sacrifice and willingness to endure the hardships and lonely moments during my absence. I am sincerely indebted to the tireless support, opportunity and prayers of my parents towards my education. May God bless you Josephine (mum), and for my dad, the late Jean Pierré who passed on while I was for this research, *eternal rest grant unto him, O Lord, and let your perpetual light shine upon him; may he rest in peace, Amen.* I appreciate the general support of Sr. Pasqua of Comboni Missionaries, Eric Kayom, my brothers and sisters, in-law family, and all the rest.

I extend a very hearty vote of thanks to the International English-speaking University Parish of KU Leuven for providing favorable atmosphere for spiritual nourishment. I also cannot forget how pleasurable it was for me to work in a stimulating research environment provided by a team of very enthusiastic and resolute members of the Hydraulics laboratory; keep up the vibrant spirit! I extend a posthumous recognition to the late Dr. Eng. Paul Nyeko-Ogiramoi (a former colleague at the Hydraulics laboratory) for his encouragement and support before he passed on, *may the good Lord grant his soul eternal rest peace, Amen.*

I am greatly indebted to Agnieszka and Martin (and their children Joanna, Mieszko, and Karol) in Poland for their continuous encouragement and readiness to always warmly welcome me into their home every time I needed a conducive place to relax my mind off the stressful research work. Thanks for the wonderful moments of the historical and heritage tours.

Charles Onyutha

English summary

Rainfall variability engenders changes in the frequency and severity of water-related disasters. In the Nile basin, rainfall-based disasters *e.g.* landslides, flooding and dry conditions are common. This research aimed at providing a platform for insights on variability of historical rainfall and flow in the River Nile basin. Spatio-temporal variability of both rainfall volumes and extreme events across the study area was analyzed in a quantile-based way. The possible drivers of the rainfall variability were investigated. The influence of spatio-temporal scales on rainfall variability analysis was investigated using the empirical orthogonal functions. The main driver of flow variability was investigated in both data- and simulation-based ways through assessment of rainfall-flow co-variation and rainfall-runoff modeling respectively. Uncertainty in sampling of hydro-climatic quantiles was assessed using extreme value analysis.

Both rainfall totals and extreme events exhibited oscillatory behavior over multi-decadal time scales across the study area including the Southern (SOU), Central (CEN) and Northern (NOR) regions. Rainfall variability possible drivers were found to be the variation in: Sea Surface Temperatures (SST) of the Atlantic and Pacific Oceans (for CEN), sea level pressure of the North Atlantic Ocean and the SST of the Indian Ocean (for SOU), and the SST of Indian and Atlantic Oceans as well as from the Pacific Ocean (for NOR). The amount of explained variability was found to reduce with the increase in temporal resolution or spatial domain size. Rainfall variability was found to be more suitably explained at a regional than location-specific spatial scale. The largest amount of river flow variation over the data period 1950-2000 was found to be explained by only the rainfall variability. The H_0 (no correlation between rainfall and flow) was rejected at the 5% significance level for several catchments. The sampling uncertainty in terms of the difference between the limits of the 95% CIs expressed as a percentage of the empirical 10-year daily quantile was, on average, 35% (for rainfall intensity), 22% (for high flow), and 51% (for low flow).

Nederlandse samenvatting

Ruimtelijke en tijdsvariabiliteit van neerslag is de belangrijkste verklarende factor voor de ruimtelijke en tijdsvariaties in het voorkomen van water-gerelateerde rampen. In het rivierbekken van de Nijl zijn er meerdere typen van zulke rampen: overstromingen, droogterampen en aardverschuivingen. Dit onderzoek had als doel om nieuwe inzichten te verwerven m.b.t. de variabiliteit van neerslag en rivierdebieten in het Nijlbekken en dit op basis van langjarige historische meetreeksen (voor de meeste stations langer dan 1950-2000). Voor neerslag werden zowel de ruimtelijke als de temporele variabiliteit bestudeerd en dit voor zowel de neerslagvolumes als de extreme neerslagintensiteiten. Voor de debieten werden zowel de piekdebieten als de laagwaterafvoeren geanalyseerd. De grootschalige atmosferische en oceanografische invloedsfactoren die aan de basis liggen van deze variaties werden in kaart gebracht. Dit gebeurde via empirische orthogonale functions. De bijhorende variaties in rivierdebieten werden via twee verschillende methoden onderzocht: via dezelfde data-gebaseerde aanpak als voor de neerslag; en via conceptuele hydrologische modelsimulaties om de neerslagvariaties om te zetten in bijhorende debietvariaties. Bij de analyse van de extremen ging bijzondere aandacht naar de onzekerheden als gevolg van de beperkte databeschikbaarheid.

De resultaten tonen aan dat zowel de neerslagtotalen als de neerslag- en debietextremen onderhevig zijn aan temporele klimaatschommelingen over tijdschalen van één of meerdere tientallen jaren. De fasering van deze klimaatschommelingen varieert evenwel tussen de zuidelijke (SOU), centrale (CEN) en noordelijke (NOR) gebieden van het Nijlbekken. Deze tijdsvariaties konden worden toegeschreven aan langjarige tijdsvariaties in de temperatuur aan het zeeoppervlak (SST) van de Atlantische en Stille Oceanen (voor CEN), de luchtdruk op zeeniveau in de Noord-Atlantische Oceaan en de SST in de Stille Oceaan (voor SOU) en de SST in zowel de Indische en Atlantische Oceanen als de Stille Oceaan (voor NOR). Het

percentage van de totale variatie dat verklaard kon worden door deze teleconnecties bleek afhankelijk van de tijdsresolutie en de grootte van het bestudeerde gebied. Hoe groter het gebied en hoe kleiner de tijdsresolutie, hoe kleiner dat percentage. Neerslagvariabiliteit wordt best geanalyseerd op de regionale schaal i.p.v. lokaal. De neerslagvariaties bleken logischerwijs ook de grootste bijdrage te leveren in het verklaren van de debietvariaties. De onzekerheidsanalyse gaf aan dat de onzekerheid op de extreme waarden met een terugkeerperiode van 10 jaar, uitgedrukt als de bandbreedte van het 95%-betrouwbaarheidsinterval, ongeveer 35% bedraagt van de extreme waarde voor neerslagintensiteiten, 22% voor piekdebieten en 51% voor laagwaterdebieten.

Acronyms and notations

| ACRONYMS | | | |
|-------------------|------------------------------------|------------|--|
| ASA | Autocorrelation Spectral Analysis | MDS | Maximum Dry Spell |
| AMO | Atlantic Multi-decadal Oscillation | MK | Mann-Kendall |
| AMS | Annual Maximum Series | ML | Maximum Likelihood |
| CI | Confidence Interval | MLR | Multiple Linear Regression |
| CRU | Climatic Research Unit | MOM | Method of Moments |
| CSD | Cumulative Sum of Difference | MSE | Mean Squared Error |
| CV | Coefficient of Variation | NAM | <i>Nedbør-Afstrømnings-Model</i> |
| ENSO | El Niño-Southern Oscillation | NAO | North Atlantic Oscillation |
| EVD | Extreme Value Distribution | NPI | North Pacific Index |
| EHRI | Extreme High Rainfall Intensity | OH | Oscillation High |
| FDF | Flow Frequency Duration | OL | Oscillation Low |
| EOF | Empirical Orthogonal Functions | PC | Principal Component |
| FAO | Food & Agriculture Organization | PDD | Percentage of Dry Days |
| FDF | Flow Duration Frequency | PDO | Pacific Decadal Oscillation |
| GCM | General Circulation Model | PDS | Partial Duration Series |
| GEV | Generalized Extreme Value | PET | Potential Evapo-Transpiration |
| GPD | Generalized Pareto Distribution | PGF | Princeton Global Forcings |
| GLUE | Generalized Likelihood | POT | Peak-Over-Threshold |
| | Uncertainty Estimation | POME | Principle of Maximum Entropy |
| HBV | <i>Hydrologiska Byråns</i> | PPCC | Probability Plot Correlation Coefficient |
| | <i>Vattenavdelning</i> | | |
| HRI ₁₀ | ten Highest Rainfall Intensities | PWMs | Probability Weighted Moments |
| IDW | Inverse Distance Weighted | QPM | Quantile Perturbation Method |
| IOD | Indian Ocean Dipole | Q-Q | Quantile-Quantile |
| ITCZ | Inter-Tropical Convergence Zone | RMSE | Root Mean Squared Error |
| JJAS | June to September | SEE | Standard Error of the Estimate |
| K-S | Kolmogorov-Smirnov test | SLP | Sea Level Pressure |
| Lmom | L-moments | SOI | Southern Oscillation index |
| LTM | Long-Term Mean | SST | Sea Surface Temperature |
| LVB | Lake Victoria Basin | TPI | Trans Polar Index |
| MAM | March to May | WLR | Weighted Linear Regression |
| NOTATIONS | | | |
| $G(x)$ | cumulative distribution function | K_{stat} | K-S test statistic |
| ξ | location parameter | K_{CV} | critical value of K_{stat} |
| k | shape parameter | E_{NS} | Nash-Sutcliffe Efficiency |
| x_t | threshold | R^2 | coefficient of determination |
| γ | extreme value index | T_{max} | daily maximum temperature |
| α | scale parameter | T_{min} | daily minimum temperature |
| α_s | level of significance | T | return period |
| H_0 | null hypothesis | H_1 | alternative hypothesis |

Table of contents

Chapter One 1

1 Introduction..... 1

1.1 Background..... 1

1.2 Problem statement..... 2

1.3 Main aim and specific objectives 3

1.3.1 Main aim..... 3

1.3.2 Specific objectives..... 3

1.4 Research questions 3

1.5 Research scope..... 4

1.5.1 Spatio-temporal scales 4

1.5.2 Climate variation..... 5

1.5.3 Climate variability versus climate change 5

1.5.4 Detection and attribution..... 6

1.6 Outline of this thesis 7

Chapter Two 9

2 Study area and data 9

2.1 Study area..... 9

2.2 Data 11

2.2.1 Rainfall 11

2.2.2 River flow..... 21

2.2.3 Potential evapotranspiration 23

2.2.4 Series related to large-scale ocean-atmosphere interactions. 24

Chapter Three 27

3 Variability in seasonal and annual rainfall 27

3.1 Introduction 27

3.2 Methodology 29

3.2.1 Computing variability using the QPM..... 29

3.2.2 Test of significance of the QPM anomalies 30

3.2.3 *Spatial differences in rainfall anomalies* 31

3.2.4 *Correlation analysis*..... 32

3.3 Results and discussion 33

3.3.1 *Differences in temporal anomalies of rainfall* 33

3.3.2 *Co-variation of the SLP and rainfall variability* 35

3.3.3 *Correlation analysis*..... 39

3.4 Conclusions and relevance of the findings 48

3.4.1 *Answers to research questions* 48

3.4.2 *Relevance of the findings* 48

Chapter Four 49

4 Variability in extreme rainfall 49

4.1 Introduction 49

4.2 Methodology 49

4.3 Results and discussion..... 50

4.3.1 *Spatial distribution in dry conditions* 50

4.3.2 *Variability in the highest rainfall intensities*..... 51

4.3.3 *Variability in the lowest rainfall intensities* 56

4.3.4 *Significance of the temporal variability*..... 63

4.4 Conclusions and relevance of the findings 65

4.4.1 *Answers to research questions* 65

4.4.2 *Relevance of the findings* 65

Chapter Five..... 67

5 Influence of spatio-temporal scales on variability analyses 67

5.1 Introduction 67

5.2 Methodology 68

5.2.1 *Extraction of data for rainfall variability analysis* 68

5.2.2 *EOF analyses of rainfall variability*..... 70

5.2.3 *Correlation analysis*..... 73

5.3 Results and discussion..... 73

5.3.1 *Explained variance versus spatial scale* 73

5.3.2 *Explained variance versus temporal scale* 74

5.3.3 *Spatial and temporal variability* 76

5.3.4 *Correlation analysis*..... 79

5.4 Conclusions and relevance of the findings 85

 5.4.1 *Answers to research questions* 85

 5.4.2 *Relevance of the findings*..... 85

Chapter Six 86

6 Flow-rainfall co-variation 86

 6.1 Introduction 86

 6.2 Selected catchments..... 89

 6.3 Methodology 90

 6.3.1 *Co-variation of flow and rainfall* 90

 6.3.2 *Rainfall-runoff modeling* 91

 6.3.3 *Analyses of the changes in the flow* 96

 6.4 Results and discussion 104

 6.4.1 *Co-variation of flow and rainfall* 104

 6.4.2 *Rainfall-runoff modeling* 105

 6.5 Conclusions and relevance of the findings 118

 6.5.1 *Answers to research questions* 118

 6.5.2 *Relevance of the findings*..... 118

Chapter Seven 119

7 Uncertainty in sampling hydro-climatic extremes 119

 7.1 Introduction 119

 7.2 Uncertainty in rainfall intensity quantiles..... 121

 7.2.1 *Methodology*..... 121

 7.2.2 *Results and discussion* 133

 7.3 Uncertainty in flow quantiles..... 142

 7.3.1 *Methodology*..... 142

 7.3.2 *Results and discussion* 144

 7.4 Conclusions and relevance of the findings 149

| | | |
|---------------------|---|-----|
| 7.4.1 | <i>Answers to research questions</i> | 149 |
| 7.4.2 | <i>Relevance of the findings</i> | 149 |
| Chapter Eight | | 150 |
| 8 | Summary and recommendations..... | 150 |
| 8.1 | Summary..... | 150 |
| 8.2 | Recommendations..... | 150 |
| 9 | Bibliography | 152 |
| 10 | Information about the author and list of some publications | 169 |

List of Figures

Figure 1.1 Outline of this dissertation8

Figure 2.1 The Nile basin and the river riparian countries..... 10

Figure 2.2 Locations of the meteorological stations 13

Figure 2.3 Long-term mean monthly rainfall 13

Figure 2.4 Locations of hydro-meteorological stations..... 15

Figure 2.5 The CV of observed and PGF-based rainfall 18

Figure 2.6 Comparison of observed and PGF rainfall..... 19

Figure 2.7 Selected catchments in the Nile Basin 21

Figure 3.1 QPM results for annual rainfall..... 34

Figure 3.2 Correlation between annual SLP and rainfall 36

Figure 3.3 Correlation between annual SST and rainfall 38

Figure 3.4 Annual SLP differences and rainfall..... 40

Figure 3.5 QPM results for annual rainfall and climate indices. 44

Figure 4.1 Spatial distribution of PDD and MDS..... 51

Figure 4.2 Correlation between the HRI₁₀ and climate indices 53

Figure 4.3 Variation in HRI₁₀ along with climate indices..... 56

Figure 4.4 Correlation between the PDD and climate indices 56

Figure 4.5 Variation of PDD along with the climate indices 59

Figure 4.6 Correlation MDS and climate indices 60

Figure 4.7 Variation in the MDS along with climate indices..... 61

Figure 4.8 Locations with significant HRI₁₀ and dry conditions 64

Figure 5.1 Spatial domains of the rainfall series 69

Figure 5.2 The River Nile riparian countries 70

Figure 5.3 Climate indices and the EOF factor scores 73

Figure 5.4 Amount of variability explained in the annual rainfall 75

Figure 5.5 Variability explained by the first principal component..... 76

Figure 5.6 Leading EOF factor loadings for rainfall of various scales 77

Figure 5.7 Leading EOF factor loadings for annual rainfall 78

Figure 5.8 Leading PC for 10°×10° size of spatial domain 79

Figure 5.9 Co-occurrence of the rainfall and climate indices..... 81

Figure 5.10 Co-variation of rainfall and selected climate indices..... 82

Figure 5.11 Correlation between climate indices and annual rainfall 83

Figure 6.1 General processes of VHM, NAM, and HBV 96

Figure 6.2 Some forms of changes identifiable using CSD approach 100

Figure 6.3 Variation in flow and rainfall 104

Figure 6.4 Simulated *versus* observed daily flow 107

Figure 6.5 Evaluation of the model performance 109

Figure 6.6 Variability in observed and modeled overland flow..... 112

Figure 6.7 Cumulative sum of variation in observed and modeled flow... 113

Figure 6.8 Cumulative sum of variation in model residuals..... 115

Figure 7.1 Observations and calibrated EVD 130

Figure 7.2 Cumulative probability distribution of the γ 134

Figure 7.3 Maximum, mean, minimum and 95% CI of the γ 135

Figure 7.4 Deviation of the EVD from the empirical quantiles..... 136

Figure 7.5 Probability distribution of the *SEE*..... 137

Figure 7.6 Parameter α from different parameter estimation methods ... 138

Figure 7.7 Performance of various parameter estimation methods..... 139

Figure 7.8 GPD parameters estimated by the WLR method..... 140

Figure 7.9 Empirical *T versus* EVD based *T*..... 141

Figure 7.10 Daily high and low flow 143

Figure 7.11 FDF relationships of high and low flow 145

Figure 7.12 Bias and RMSE on return periods..... 146

Figure 7.13 Uncertainty on return periods 147

Figure 7.14 Uncertainty on 10-year quantiles 148

List of Tables

Table 2.1 Overview of selected stations and some statistics 14

Table 2.2 Selected stations with daily rainfall data..... 16

Table 2.3 Correlation between observed and PGF rainfall 18

Table 2.4 Monthly rainfall stations for selected watersheds 20

Table 2.5 Selected hydrological stations and flow data 22

Table 2.6 Overview of data at the selected hydrological stations..... 23

Table 2.7 Overview of SLP, SST and related time series 26

Table 3.1 Correlation between rainfall and SLP differences 41

Table 3.2 Correlation between annual rainfall and climate indices..... 45

Table 3.3 Correlation between seasonal rainfall and climate indices 47

Table 4.1 Adjusted R^2 in predicting HRI_{10} using climate indices 55

Table 4.2 Adjusted R^2 in predicting PDD using climate indices 58

Table 4.3 Adjusted R^2 in predicting MDS using climate indices..... 63

Table 5.1 Correlation between rainfall variability and climate indices..... 84

Table 6.1 Correlation between the variation of rainfall and river flow 105

Table 6.2 Observed and simulated daily flow..... 107

Table 6.3 List of model parameters considered for calibration..... 110

Table 6.4 The p -values of MK and CSD trend tests 114

Chapter One

Introduction

1.1 Background

Apart from the historical legacy of restive trans-boundary water management issues, the Nile basin faces many other formidable challenges *e.g.* floods, droughts, food insecurity resulting from combined effects of rainfall variability and the unswerving dependence of majority of the population on subsistence supported by rain-fed cropping system, *etc.* In the Nile basin, the problems of both dry conditions and water surplus exist.

Extremely dry conditions tend to occur in several parts of the study area *e.g.* Karamoja region in the North Eastern Uganda, Megenta area of Afar in Ethiopia, the Sahel region, Egypt, *etc.* Some of the past devastating drought events include that of: the early 1980s in the Ethiopian Highlands (Sutcliffe *et al.*, 1989), the 1970s and 1980s which led to the decline in annual rainfall by 30% in the Sahel region (Hulme, 1992), *etc.* Prolonged dry conditions (*i.e.* sustained insufficiency of water availability) causes negative environmental, social and economical impacts.

The occurrence of heavy rainfall events seems common especially in the Equatorial region and Ethiopia. Some of the recent disasters which claimed lives and property in these areas include: 1) the devastating flooding in Kasese district of Uganda in the early May 2013 and mid-May 2016, 2) the deadly landslides at the foot of Mount Elgon slope at the Uganda-Kenya border which occurred in March 2010 and June 2012, and 3) the disastrous floods and landslides in Oromia and Southern Nations, Nationalities and People's Region of Ethiopia in May 2016. Some typical past rainfall extremes reported for the study area include: daily rainfall of 88 mm (which was the highest since late 1890s) in Khartoum on the 31st July 1920 (Hulme and Trilsbach, 1989), the severe floods of August and September in 1988 in Sudan (Sutcliffe *et al.*, 1989), *etc.* In the history of flooding in the Nile basin, Ethiopia, Tanzania, Kenya, Sudan and Uganda are the most affected countries in terms of the average number of flooding occurrences

(Kibiiy *et al.*, 2010). Actually, for the Nile basin countries in the Equatorial region (*e.g.* those which share the Lake Victoria Basin (LVB)), flooding is one of the most devastating natural hazards as reported by the United Nations Environmental Programme UNEP (2006). The disasters (*e.g.* landslides, floods) which occur due to Extreme High Rainfall intensities (EHRI) lead to loss of lives and property.

1.2 Problem statement

The problems of both EHRI and dry conditions identified in Section 1.1 pose serious challenges to the livelihoods and complicate water resources planning and management in the study area. Information on the historical or observed variability of flow or rainfall and associated possible drivers is vital to develop measures for adaption to the effects of climate variability on extreme hydro-climatic conditions. To ensure the information on past climate variation is relevant for multipurpose applications (*e.g.* hydro-electric power generation, domestic water supply, irrigation, *etc.*), analyses of rainfall or flow variability should, as done in this study,

- a) cover the entire River Nile basin,
- b) be based on various temporal scales *e.g.* daily, seasonal, annual, *etc.*,
- c) incorporate uncertainty on the hydro-climatic quantiles, and
- d) comprise the assessment of factors which influence variability results.

Although several studies were conducted on variability of rainfall or flow in the River Nile basin, they did not take in account the above listed considerations a) to d). Studies by *e.g.* Berhane *et al.* (2014), Taye and Willems (2012), Nyeko-Ogiramoi *et al.* (2013), and Phillips and McIntyre (2000) did not consider the entire River Nile basin but were confined to sub-basins or sub-regions. Other studies such as Berhane *et al.* (2014), Block and Rajagopalan (2007), and Abtew *et al.* (2009) focused on rainfall totals but not the extreme rainfall intensities. Occasional studies which analyzed the variability in daily hydro-climatic quantiles (see *e.g.* Taye and Willems, 2012; Nyeko-Ogiramoi *et al.*, 2013) did not quantify the uncertainty due to sampling of the extreme events. Several studies which

investigated flow changes using hydrological models (see *e.g.* Mango *et al.*, 2011; Olang and Fürst, 2011; Legesse *et al.*, 2003; 2004, Gebrehiwot *et al.*, 2013; Taye and Willems, 2013) lacked insight on the influence of model selection on simulation results.

The need for analyses which could address the above gaps on information about flow or rainfall changes from previous studies, given the relevance of climate variability for water resources management, compelled the undertaking of this research.

1.3 Main aim and specific objectives

1.3.1 Main aim

The main aim of the research was to explore and present a platform for understanding the temporal variability in rainfall and flow in support of water resources management and decision making at regional and local (sub-basin) scales in the Nile basin.

1.3.2 Specific objectives

The specific objectives were:

- 1) To analyze spatio-temporal variability of rainfall while investigating possible linkage to large-scale ocean-atmosphere interactions,
- 2) To assess the variability in EHRI and dry conditions,
- 3) To determine the influence of spatial and temporal scales on the rainfall variability analyses,
- 4) To determine the extent to which the flow variability can be explained by rainfall variation, and
- 5) To quantify uncertainty in sampling of hydro-climatic extreme events.

1.4 Research questions

For each of the five pathways advanced towards achieving the main aim of this research, answers to several questions were sought. The research questions were grouped according to each specific objective.

- a) For *specific objective 1*: Is there variability in rainfall? Is the variability of a particular pattern? Is the variability significant? Is the variability linked to large-scale ocean-atmosphere interactions?
- b) For *specific objective 2*: Is there variability in extreme rainfall conditions? Is the variability of dry conditions and EHRI significant? Can the variability of the extreme rainfall be explained in terms of the variation in large-scale ocean-atmosphere conditions?
- c) For *specific objective 3*: How does selection of temporal and spatial scales influence results of rainfall variability analyses? Does the choice of the spatio-temporal scales affect the linkage of rainfall variability to large-scale ocean-atmosphere interactions?
- d) For *specific objective 4*: Can the variability in daily overland flow be adequately explained by the variation in daily rainfall? Was there a change in the response of catchment (as a system) to hydro-climatological inputs?
- e) For *specific objective 5*: What is the best class of the Extreme Value Distribution (EVD) to model extreme hydro-climatic events of the study area? What is the best method for estimation of the EVD parameters? What is the sampling uncertainty on rainfall and river flow quantile estimates?

1.5 Research scope

1.5.1 Spatio-temporal scales

For analyses of rainfall variability, regional scale was considered. Furthermore, for variability analyses, coarse temporal scales such as annual, seasonal and monthly resolutions were used. Coarser temporal scale, *e.g.* annual resolution, presents better overview of the possible driving forces of the rainfall variability than that of the finer (*e.g.* daily and hourly) time scales.

Frequency analyses of hydro-climatic extreme events were based on daily series. Analysis using daily data was limited to the LVB as the wettest part of the study area.

1.5.2 Climate variation

Climate variation has various aspects that can be analyzed. The choice of which climate variation aspect to consider for analysis depends on the purpose of the study.

It is known that climate keeps varying. According to The Energy and Resources Institute (TERI, 2017), climate variation can be natural or human-induced. Human-induced climate variation can be as a result of: i) the changes in vegetation and land-uses which impact on regional surface energy and water balance, ii) greenhouse gases from thermal plants, petrol or diesel vehicles, *etc.* Natural climate variation can result from forcings or variability. Natural variability of climate can be i) externally induced *e.g.* by the changes in solar radiation, or ii) internally induced *e.g.* by the interactions between oceans and atmospheres, biosphere and atmosphere. Still under natural climate variation, forcings can be from i) an external source *i.e.* linked to intensity of the solar radiation, fluctuations in solar energy, or ii) internal *e.g.* fluctuations in ocean circulations, large-scale changes in terrestrial biosphere, *etc.*

The scope of this research was limited to internally induced climate variability. The main driving force of the climate variability considered was the variation in the large-scale ocean-atmosphere conditions. Due to its direct relevance for water management applications, rainfall was the main weather element used for investigating the spatio-temporal influence of the climate system on hydrology of the study area.

1.5.3 Climate variability versus climate change

To properly understand the results of this research, it is important to note that climate variability (and not climate change) was considered. Climate change requires the consideration of long-term periods, preferably 30 years and above. Furthermore, climate change analysis tends to focus on the context of whether there is long-term trend or not. Climate variability comprises the use of decades for analysis. When periods of *e.g.* 10 or 15 years are considered, as in this study, negative or positive trend (which can

be opposite in directional sign to that when, say, 200 years are used) may be obtained. This means, possible attempts to relate such short-durational trend results to climate change are unreliable and can be highly misleading.

1.5.4 Detection and attribution

Grammatically, to detect means identify the presence or existence of something, and to attribute means consider something resulting due to a cause. However, when it comes to the science of climate change, the words detection and attribution are technical in what they mean. Detection does not imply attribution of the detected change to the assumed cause (Intergovernmental Panel on Climate Change (IPCC), 2001). According to the IPCC (2001), the identification of change in observations can be based on the expected responses to external forcing, and this can be from the simulation by climate models or physical understanding. An identified change is said to be '*detected*' if its likelihood of occurrence by chance due to internal variability alone is verified to be small *i.e.* with respect to the defined level of confidence (IPCC, 2001). Ideally,

".....unequivocal attribution would require controlled experimentation with the climate system. Since that is not possible, in practice attribution of anthropogenic climate change is understood to mean demonstration that a detected change is consistent with the estimated responses to the given combination of anthropogenic and natural forcing and not consistent with alternative, physically plausible explanations of recent climate change that exclude important elements of the given combination of forcings" (IPCC, 2001; 2007).

In this study, co-occurrence of independent and dependent variables was investigated using Pearson correlation test and may indicate statistical measure of association. The correlation analyses of this research were not done in the context of practical attribution of anthropogenic climate change. In other words, no climate models were used for simulation and neither was controlled experimentation with the climate system done. The correlative relationship between the independent (*e.g.* climate indices) and dependent (*e.g.* rainfall) variables, amidst data scarcity as for the study

area, can be useful to obtain improved quantile estimates for short-term planning of water resources applications.

1.6 Outline of this thesis

In *Chapter two*, the study area and various data considered for the proposed pathways are described. For each specific objective, the particular dataset(s) used in answering the research questions is (are) clearly mentioned. How the hydro-climatic data was used for other research components is summarized in Figure 1.1.

Chapter three describes how the first specific objective (*i.e.* variability of rainfall volumes and the linkage to ocean-atmosphere interactions) was undertaken. Methodology, results, discussion as well as the conclusions are also presented.

Chapter four comprises how the questions regarding variability of rainfall extreme conditions and the possible co-occurrence with climate indices were answered.

In *Chapter five*, how the selection of spatial and temporal scales influences the analysis of rainfall variability is tackled.

In *Chapter six*, investigation was made on whether, for the data period considered, the amount of flow variability that could not be explained by rainfall variation was large. Results on whether the possible change in catchment characteristics impacted significantly on the rainfall-runoff generation process are also presented and discussed.

Given that the results of the variability analyses are in quantile-based form, the assessment of the uncertainty in sampling of extreme rainfall and flow events are presented in *Chapter seven*.

Chapter eight gives the new insights from the study relevant for water resources management. In particular, the answers to the various research questions of the study are provided. Recommendations for further or future research are also given.

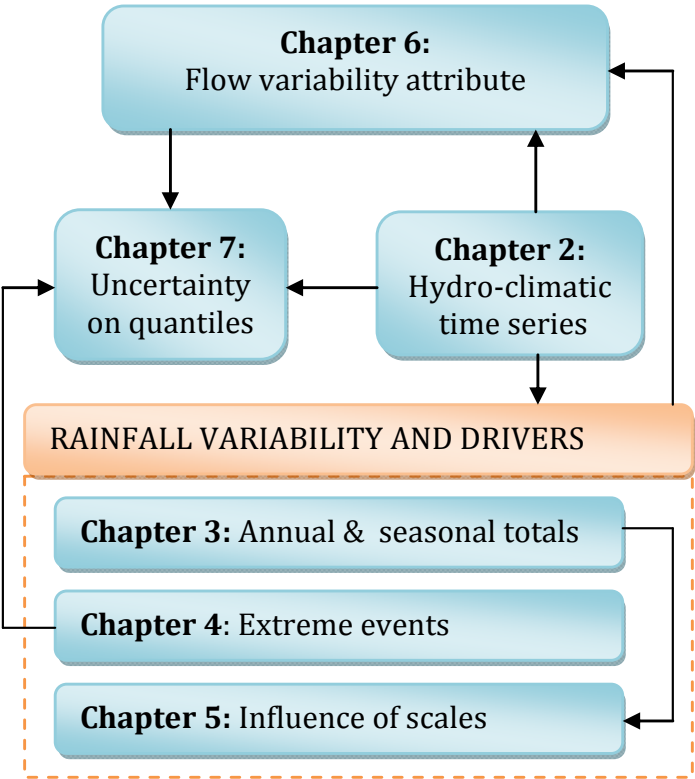


Figure 1.1 Outline of this dissertation

Chapter Two

Study area and data

2.1 Study area

The Nile basin, which has a total drainage area of about 3,400,000 km² (see Figure 2.1), stretches over 35° of latitude in North–South direction (31° N to 4° S) and over 16° of longitude in West–East direction (24 to 40° E). It comprises the River Nile which is 6,695 km from its furthest source (Ruvyironza in Kagera basin) to the Nile Delta in Egypt. River Nile is the world's longest river under arid condition and is fed by two main river systems: the White Nile (from the Equatorial region), and the Blue Nile (from the Ethiopian Highlands). The major lakes within the River Nile basin include Lake Victoria, Edward, Kyoga, Albert, No, Tana, and Nasser. According to Food and Agriculture Organization (FAO, 1997), the percentage of the River Nile riparian countries covered by the Nile basin includes: Uganda (98.1%), Kenya (8.0%), Tanzania (8.9%), Rwanda (75.5%), Burundi (47.6%), Democratic Republic of Congo (0.9%), Sudan and South Sudan (79.0%), Ethiopia (33.2%), Eritrea (20.4%), and Egypt (32.6%). The River Nile receives little or negligible runoff from about 40% of its drainage (basin) area which comprises dry lands of either arid or hyper-arid condition. Moreover, the River Nile loses about 50% of its water by evaporation in the floodplains of the Sudd region of South Sudan (FAO, 2016). The impacts of these factors result in low runoff efficiency in the Nile basin despite its vast drainage area.

The majority of the people in the Nile basin depend on rain-fed agriculture to support livelihoods. However, the climate of the Nile basin is characterized by a strong latitudinal wetness gradient (Camberlin, 2009). There is a spatially contrasted distribution of mean annual rainfall over the Nile basin. About 28% of the basin receives less than 100 mm annually. Whereas some parts experience hyper-arid condition, substantial area exhibits sub-humid condition. Rainfall in excess of 1000 mm is restricted mainly to the Equatorial region and the Ethiopian Highlands. From Northern Sudan all across Egypt, rainfall is negligible (below 50 mm except

along the Mediterranean coast). This general distribution reflects the latitudinal movement of the Inter-Tropical Convergence Zone (ITCZ) which never reaches Egypt and Northern most Sudan, stays only briefly in Central Sudan and longer further South (Camberlin, 2009).

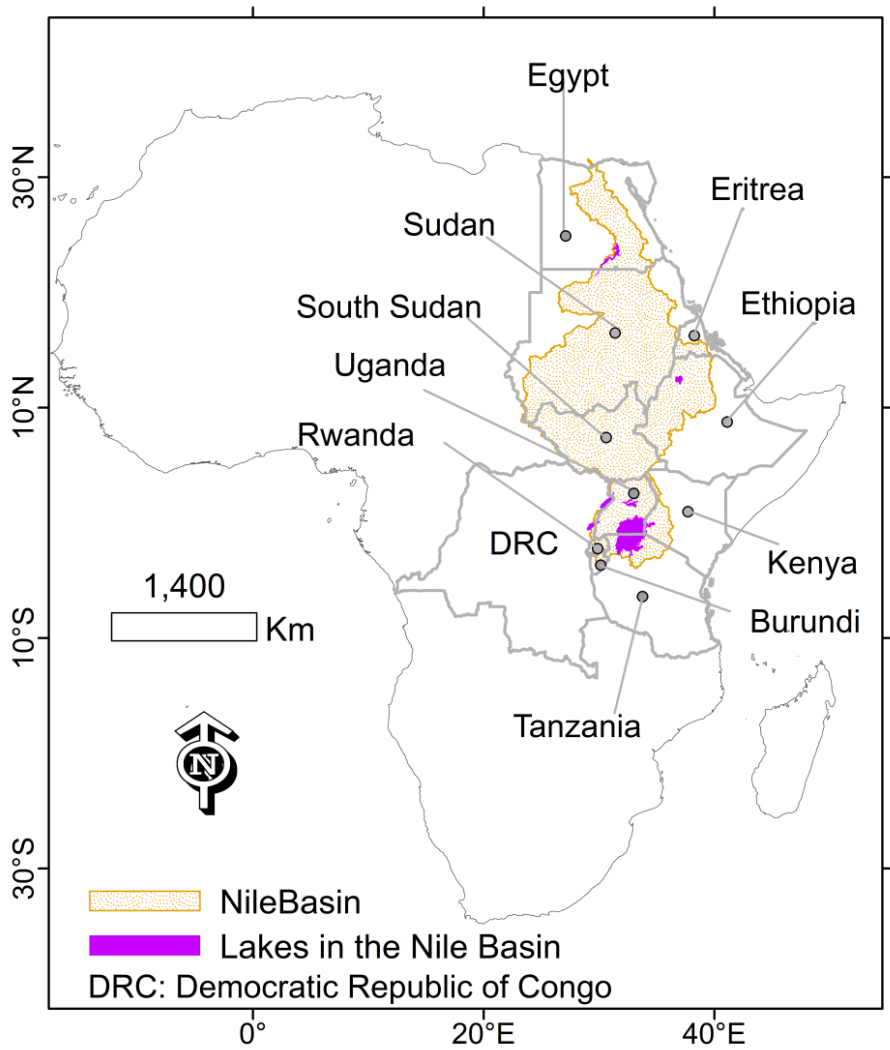


Figure 2.1 The Nile basin and the river riparian countries

2.2 Data

Hydro-meteorological series were obtained from various sources.

2.2.1 Rainfall

a) Series for study on seasonal and annual rainfall variability

The study on seasonal and annual rainfall variability presented in Chapter 3 was based on monthly data obtained from the FAO (2001) agro-climatic database incorporated in the FAOCLIM 2 tool downloaded via <http://www.fao.org/> (accessed: January 2010). FAOCLIM contains monthly world-wide agro-climatic data for 28,100 stations with up to 14 observed and computed agro-climatic variables. To enhance the acceptability of the research findings, only rainfall series recorded over 40 years with missing data points less than 10% of the full sample size (considering the entire record period) were used. Missing data points were filled-in using the Inverse Distance Weighted (IDW) interpolation technique of Shepard (1968) (Eq. 2.1). The missing rainfall intensity (P_Ω) at Station Ω for a certain period using rainfall values (P_j) at g neighboring stations for the same period is given by:

$$P_\Omega = \sum_{j=1}^g P_j \times d_j^{-r} \left(\sum_{j=1}^g d_j^{-r} \right)^{-1} \quad (2.1)$$

where d_j is the distance between Station Ω (the one with the missing record) and that in the neighborhood being used for interpolation, and r is the power parameter (an arbitrary positive real number).

It is important to note that the reliability of the IDW interpolation lies in the value of r . A small value of r tends to approximate the interpolated record to average of the values from the neighboring stations used, while a large value of r down-weights points further away and assigns large weights to the nearest points (Lu and Wong, 2008). According to Dirks *et al.* (1998), interpolation errors appear to be minimized by taking $r = 2$ for daily or monthly data, 3 for hourly and 1 for annual data. Hence $r = 2$ was adopted in this study. Due to the difference in rainfall statistics across the study

area, the IDW was applied taking proximity of the stations into perspective *i.e.* using stations of a particular region. To enhance the reliability of the in-filled data, a dense network of about 200 stations over all the River Nile riparian countries was used. The term g from Eq. (2.1) was set to at least 4 stations during each interpolation procedure.

Table 2.1 shows the station name, record length, Long-Term Mean (LTM, mm/year), and Coefficient of Variation (CV) of the annual rainfall time series. The CV was found to range from 0.11 (Station 30) to 1.95 (station 35) which shows that there was moderate variability on a year to year basis in the study area over the data period. The LTM of the annual rainfall was shown to vary from 1.2 mm/year (Station 34) to 2163 mm/year (Station 32). For Stations 34–35 and 37 found in the Northernmost part, the values of the CV were higher than for other regions. The high values of the CV showed that the data at these stations, characterized by large extents of ties caused by the zeros in the data, were not of Gaussian distribution. Though not shown in Table 2.1 (for brevity), the coefficient of skewness and actual excess kurtosis considering all the stations were on average 0.85 and 2.24 respectively. Thus, the data were slightly positively skewed and leptokurtic. The central peak of a leptokurtic distribution is higher and sharper and its tails are longer and fatter than that of the normal distribution.

The stations were grouped into three (see Figure 2.2) based on the long-term mean monthly rainfall pattern (Figure 2.3). The rainfall at stations in the Equatorial region (group A) is shown to exhibit bimodal pattern (Figure 2.3a) with the main wet season in March to May (MAM) and “short rains” in October to December (OND) (Nicholson, 1996). The main wet season over Sudan and Ethiopia (group B) occurs in the months of June to September (JJAS) (Figure 2.3b). For stations in Egypt (group C), it is seen that the long-term mean monthly rainfall values are far lower (and of more unclear pattern) than those in groups A and B (Figure 2.3c). However, it can be seen that the wet seasons cover MAM and the period of October to February (ONDJF).

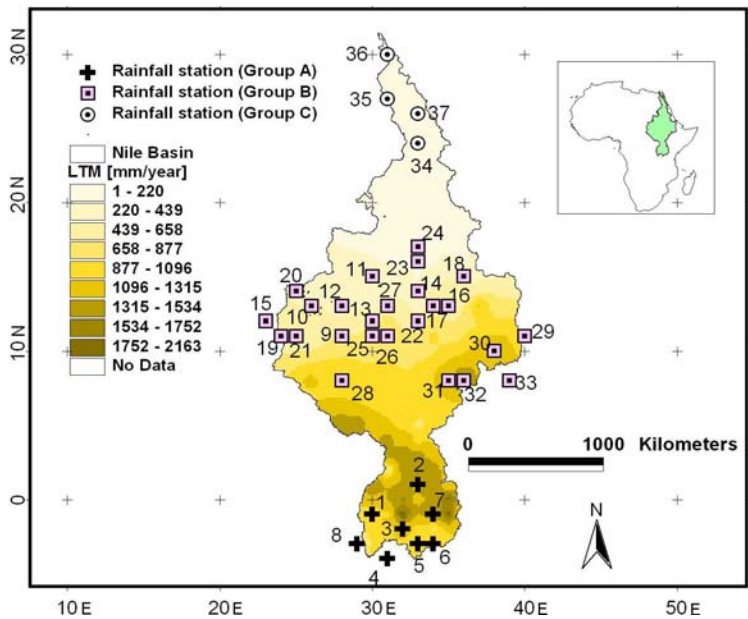


Figure 2.2 Locations of the selected meteorological stations (see Table 2.1 for details) in the Nile basin; the background is based on the annual LTM and the axes are in degrees

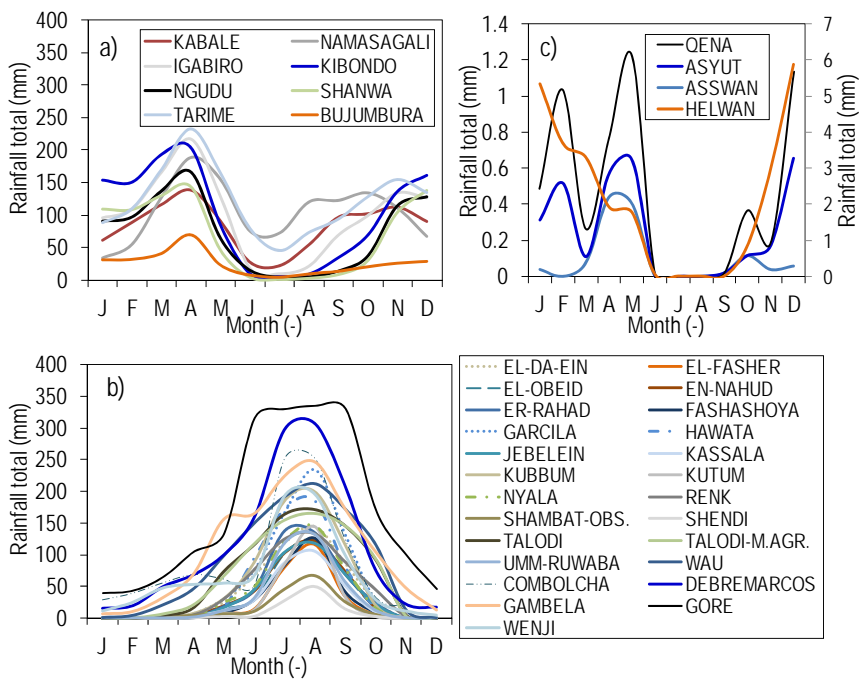


Figure 2.3 Long-term mean monthly rainfall for group a) A, b) B, and c) C

Table 2.1 Overview of selected stations and some statistics of the annual rainfall data including the CV (-) and LTM (mm/year)

| SNo. | Station | From | To | CV | LTM | SNo. | Station | From | To | CV | LTM |
|-----------------------|------------|------------------------------|------|------|------|---------------------|---------------|------|------|------|-------|
| 1 | Kabale | 1917 | 1993 | 0.17 | 1003 | 20 | Kutum | 1929 | 1990 | 0.4 | 288.1 |
| 2 | Namasagali | 1915 | 1978 | 0.19 | 1256 | 21 | Nyala | 1920 | 1996 | 0.29 | 439.5 |
| 3 | Igabi | 1931 | 1982 | 0.24 | 1198 | 22 | Renk | 1906 | 1987 | 0.2 | 507.5 |
| 4 | Kibondo | 1926 | 1978 | 0.29 | 1209 | 23 | Shambat-Obs. | 1913 | 1993 | 0.94 | 166.5 |
| 5 | Ngudu | 1928 | 1971 | 0.33 | 870 | 24 | Shendi | 1937 | 1990 | 0.74 | 100.1 |
| 6 | Shanwa | 1931 | 1985 | 0.24 | 829 | 25 | Talodi | 1916 | 1987 | 0.23 | 791.5 |
| 7 | Tarime | 1933 | 1975 | 0.2 | 1475 | 26 | Talodi-M-Agr. | 1942 | 1985 | 0.21 | 774.2 |
| 8 | Bujumbura | 1930 | 2004 | 0.18 | 251 | 27 | Umm-Ruwaba | 1912 | 1989 | 0.32 | 365.3 |
| 9 | El-Da-Ein | 1943 | 1990 | 0.27 | 474 | 28 | Wau | 1904 | 1990 | 0.16 | 1110 |
| 10 | El-Fasher | 1917 | 1996 | 0.43 | 261 | 29 | Combolcha | 1952 | 1996 | 0.17 | 1032 |
| 11 | El-Obeid | 1902 | 1996 | 0.32 | 356 | 30 | Debremarcos | 1954 | 1998 | 0.11 | 1335 |
| 12 | En-Nahud | 1911 | 1996 | 0.28 | 384 | 31 | Gambela | 1905 | 1993 | 0.22 | 1245 |
| 13 | Er-Rahad | 1931 | 1984 | 0.31 | 432 | 32 | Gore | 1946 | 1996 | 0.2 | 2163 |
| 14 | Fashashoya | 1946 | 1988 | 0.32 | 297 | 33 | Wenji | 1951 | 1994 | 0.28 | 786.9 |
| 15 | Garcila | 1943 | 1986 | 0.3 | 672 | 34 | Asswan | 1935 | 1986 | 1.92 | 1.2 |
| 16 | Hawata | 1941 | 1988 | 0.23 | 571 | 35 | Asyut | 1925 | 1986 | 1.95 | 1.9 |
| 17 | Jebelein | 1927 | 1988 | 0.29 | 377 | 36 | Helwan | 1904 | 1988 | 0.72 | 25.7 |
| 18 | Kassala | 1901 | 1996 | 0.3 | 298 | 37 | Qena | 1935 | 1986 | 1.78 | 2.4 |
| 19 | Kubbum | 1943 | 1985 | 0.27 | 646 | | | | | | |
| SNo. : Station number | | CV: coefficient of variation | | | | LTM: Long-term mean | | | | | |

b) Series for frequency analysis

To assess the uncertainty in quantile estimation, station-based daily rainfall series at 9 locations in the LVB (Figure 2.4) of the Equatorial region were obtained from the database of the FRIEND/Nile (River Nile basin Flow Regimes from International, Experimental and Network Data) project. The daily rainfall data at each station was from 1961 to 1990. Table 2.2 shows that the CV ranges from 1.4 (Station 2) to 2.7 (Station 8) thereby indicating considerable variability on a day to day basis in the LVB. The use of data at flow stations shown in Figure 2.4 can be found described in Section 2.2.2b.

To select the best parameter estimation method to be used for estimating uncertainty in quantiles at the selected rainfall stations, large series were required. Several simulation runs of various General Circulation Models (GCMs) were used as the readily available synthetic rainfall series over the LVB in order to compliment the station data and to minimize, on the study findings, the effect of the data quality problems at the selected stations. The data from the GCMs were from 1961 to 1990, a period which coincides with

that of the observed rainfall series and is also a widely used baseline for climate studies. The GCM data are also commonly applied on the basis of climate change impact studies, which is one of the important applications of the extreme quantile estimation dealt with in this study. From phase 3 of the coupled model inter-comparison project, a total of 1134 daily rainfall series were obtained (collectively for all the rainfall stations) from the archived database at program for climate model diagnosis and inter-comparison. The data were obtained online via the link http://www2-pcmdi.llnl.gov/esg_data_portal (accessed: June 2011). Some of the GCMs used were GISS-AOM, CGCM3.1(T63), MIROC3.2(hires), CSIRO-MK3.0, *etc.* The full list and the details of the GCMs used can be found in Onyutha *et al.* (2016).

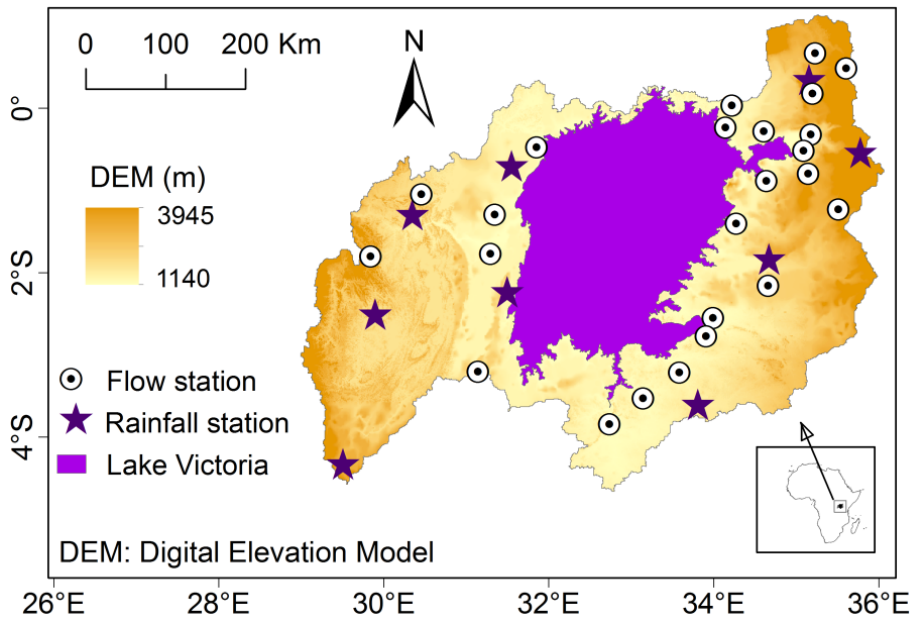


Figure 2.4 Locations of meteorological and hydrological stations (see Table 2.2, and Table 2.6, respectively for details); the background map is based on the Digital Elevation Model (DEM) obtained online from the International Centre for Tropical Agriculture, Shuttle Radar Topography Mission, via <http://strm.csi.cgiar.org/> [accessed: November 2010]

Table 2.2 Selected stations with daily rainfall data from 1961 to 1990

| SNo | Station name | Longitude [°] | Latitude [°] | Mean [mm/day] | CV |
|-----|------------------------|---------------|--------------|---------------|-----|
| 1 | Kamenyamigo | 31.67 | -0.30 | 2.6 | 2.6 |
| 2 | Kigali Aero Nyabarongo | 30.13 | -1.97 | 6.6 | 1.4 |
| 3 | Londiani Forest | 35.6 | -0.15 | 3.2 | 2.2 |
| 4 | Maswa Hydromet | 33.77 | -3.17 | 2.5 | 3.1 |
| 5 | Nyabassi | 34.57 | -1.35 | 3.9 | 2.3 |
| 6 | Rubya Mission | 31.62 | -1.72 | 3.8 | 2.3 |
| 7 | Ruvironza | 29.77 | -3.82 | 3.6 | 2.0 |
| 8 | Rwoho Forest | 30.55 | -0.85 | 2.6 | 2.7 |
| 9 | Turbo Forest | 35.02 | 0.67 | 3.6 | 2.2 |

c) Series to study the influence of spatio-temporal scales on variability

To assess the influence of spatio-temporal scales on variability analyses as presented in Chapter 5, global gridded ($0.5^\circ \times 0.5^\circ$) daily rainfall data of the Princeton Global Forcings (PGFs) (Sheffield *et al.*, 2006) were obtained online at <http://hydrology.princeton.edu/data/pgf/0.5deg/> (accessed: February 2016). The PGFs which cover the period 1948-2010 are observational-reanalysis hybrid *i.e.* derived by reanalysis dataset from the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) (Kalnay *et al.*, 1996) with a number of observational-based products from the Climatic Research Unit (CRU) (*i.e.* Time-Series (TS) Version 2.0 or CRU TS2.0), the Global Precipitation Climatology Project (GPCP), the Tropical Rainfall Measuring Mission (TRMM), and the National Aeronautics and Space Administration (NASA) Langley Research Center (see Sheffield *et al.* (2006) for the details of the data description).

To check the validity of PGF series, observed daily series from a total of 10 stations (Table 2.3) were used. Daily rainfall data at 4 locations (see stations 6-10) were obtained online from Global Historical Climatology Network (GHCN) (Menne *et al.*, 2012a; 2012b). Daily data at Station 5 was obtained from Ministry of Water and Environment, Uganda. Data at Stations 1-4 are similar to those presented in Table 2.2. At each of the selected 10 locations, daily PGF rainfall series were also extracted. Validity of the PGF rainfall was checked in two ways. Firstly, the correlation between the observed or station-based rainfall and PGF series was

analyzed at monthly, seasonal, and annual time scales. The main rainy seasons including March to May (MAM) for the Equatorial region, and June to September (JJAS) for the Ethiopian Highlands were considered for the data quality evaluation. Secondly, the CV or measure of variability in the observed series was compared with that of the PGF data. The null hypothesis H_0 of no correlation between the observed and PGF rainfall was generally rejected at significance level α_s of 5% for the various time scales (Table 2.3). For the Equatorial region (see Stations 1 to 5), aggregation of rainfall from monthly to seasonal time scale lowered the correlation. This could be thought of in terms of the differences between point rainfall when compared with the gridded rainfall averaged over some spatial domain. However, for the other Stations 6 to 10 which are from Ethiopia, it is noted that the correlation coefficients were not so much affected by the aggregation of rainfall from monthly to seasonal time scale. Therefore, the low correlation for the seasonal time scale for the Stations 1 to 5 could be because of failure of the PGF to accurately capture the influences of complex topography (with many mountains *i.e.* Mount Elgon, Mount Kenya, Mount Kilimanjaro, Mount Rwenzori, *etc*) and the Great Lakes (Victoria, Albert, *etc*) on the rainfall variation in the Equatorial region.

As the temporal resolution of the rainfall series was increased *i.e.* from annual through seasonal to monthly time scale, the CV also increased for both the observed and PGF rainfall series (Figure 2.5). The scatter points are shown to fall along the bisector. In other words, the PGF series reproduced the variability of the observed rainfall fairly well for the various time scales. Therefore, based on results shown in Table 2.2 and Figure 2.5, the quality of the PGF data was deemed adequate for the purpose of this study.

Table 2.3 Correlation between daily PGF series and observed rainfall at selected stations

| SNo. | Station name | Period | | Location | | Correlation for various time scales | | | |
|---|-----------------|--|------|----------|-------|-------------------------------------|-------------|-------------|--------|
| | | From | To | Long. | Lat. | Monthly | MAM | JJAS | Annual |
| 1 | Kamenyamigo | 1961 | 1990 | 31.67 | -0.30 | 0.62 | 0.21 | 0.15 | 0.38 |
| 2 | Maswa Hydromet. | 1961 | 1990 | 33.77 | -3.17 | 0.89 | 0.74 | 0.60 | 0.64 |
| 3 | Nyabassi | 1961 | 1990 | 34.57 | -1.35 | 0.64 | 0.60 | 0.04 | 0.53 |
| 4 | Rwoho Forest | 1961 | 1990 | 30.55 | -0.85 | 0.61 | 0.40 | 0.27 | 0.37 |
| 5 | Mbarara Met. | 1948 | 2008 | 30.6 | 0.56 | 0.59 | 0.06 | 0.47 | 0.31 |
| 6 | Bahr Dar | 1964 | 2004 | 37.42 | 11.60 | 0.91 | 0.68 | 0.44 | 0.51 |
| 7 | Debre Marcos | 1964 | 2004 | 37.67 | 10.33 | 0.96 | 0.85 | 0.67 | 0.73 |
| 8 | Gonder | 1964 | 2004 | 37.42 | 12.55 | 0.94 | 0.76 | 0.77 | 0.77 |
| 9 | Addis Ababa | 1964 | 2004 | 38.75 | 9.03 | 0.96 | 0.87 | 0.69 | 0.71 |
| 10 | Kombolcha | 1964 | 2004 | 39.73 | 11.12 | 0.91 | 0.75 | 0.79 | 0.75 |
| | | Critical values of the correlation at the significance level of 5% | | | | | | | |
| Bold values indicate H_0 (no correlation) was accepted at the significance level of 5%. | | Rainfall at Stations 1 to 4 | | | | 0.09 | 0.35 | 0.35 | 0.35 |
| | | Rainfall at Station 5 | | | | 0.07 | 0.24 | 0.24 | 0.24 |
| | | Rainfall at Stations 6 to 10 | | | | 0.08 | 0.31 | 0.31 | 0.31 |

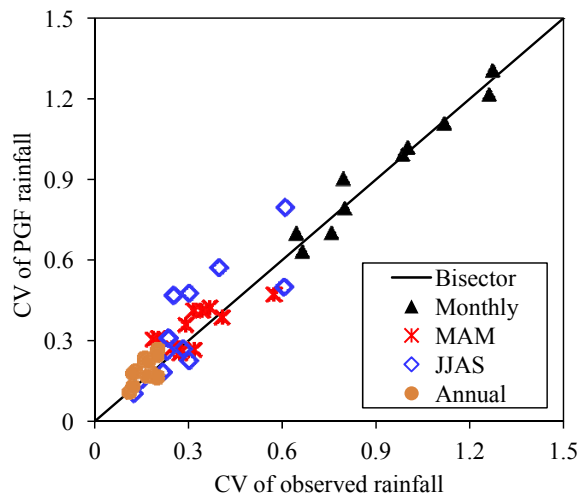


Figure 2.5 The CV of observed rainfall *versus* that of PGF gridded series

d) Series for analysis of extreme rainfall variability

The study on extreme rainfall variability presented in Chapter 4 was also based on the PGF daily gridded data. Across the study area, observed daily rainfall intensities were noted to be over- or under-estimated by the PGF-based series. This can be verified from the illustration in Figure 2.6a using the double scales for clarity of the difference between the PGF and

observed rainfall. However, the variation in some metrics characterizing extreme rainfall conditions including the longest dry spell (Figure 2.6b) based on the threshold of 1 mm/day and the mean of the 10 highest events (Figure 2.6c) in each year of the observed and PGF series was found somewhat comparable. For instance, based on Pearson correlation test, the H_0 (no correlation between the dry spell of observed and PGF series shown in Figure 2.6b) was rejected at Stations 2 and 6-10. By considering the mean of the highest events, the over- estimation of the observed EHRI also reduced tremendously (Figure 2.6c). Eventually, the PGF series were deemed suitable to analyze spatio-temporal variation of the metrics characterizing EHRI and dry conditions across the study area.

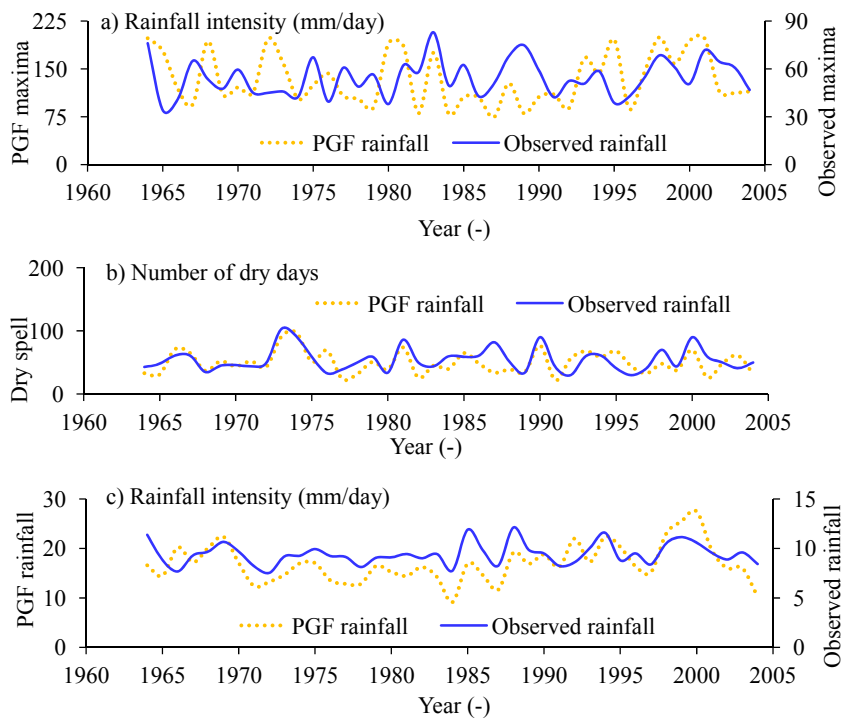


Figure 2.6 Comparison of observed and PGF rainfall (1964-2004) at Kombolcha in Ethiopia in terms of the a) maximum rainfall event in each year (mm/day), b) longest dry spell in each year, c) mean of 10 highest daily rainfall events in each year; the charts a) and c) were plotted on double scales for clarity due to the difference in the orders of magnitudes of the rainfall intensities

e) Data for the assessment of flow-rainfall co-variation

The study on flow-rainfall co-variation presented in Chapter 6 was also based on a number of rainfall stations (Table 2.4). The daily rainfall data at 5 locations over the Blue Nile were those used in a study by Taye and Willems (2013). Daily rainfall at 4 locations over the Kagera catchment, and monthly rainfall series at 6 stations in and around Atbara catchment were obtained from the FRIEND/Nile project. Furthermore, daily rainfall data from 4 stations in the Kyoga basin were obtained from the Ministry of Water and Environment, Uganda. Locations for the selected stations (see Table 2.4) are shown in Figure 2.7.

Table 2.4 Monthly rainfall stations for selected watersheds

| Study ID | Station | Location | | Data period | | Metric | |
|-----------------------|----------------|---------------------|-------|--------------------|------|--------|-------|
| | Name | Long. | Lat. | From | To | CV | LTM |
| Kagera | | | | | | | |
| Kag1 | Mugera | 29.97 | -3.32 | 1940 | 1990 | 0.78 | 102 |
| Kag2 | Muyinga | 30.35 | -2.85 | 1940 | 1992 | 0.72 | 93.6 |
| Kag3 | Igabiro Estate | 31.55 | -1.82 | 1940 | 1994 | 0.8 | 92.98 |
| Kag4 | Musenyei | 30.03 | -2.97 | 1940 | 1994 | 0.83 | 106.4 |
| Atbara | | | | | | | |
| Atb1 | Atbara | 33.97 | 17.7 | 1907 | 1995 | 3.06 | 5.93 |
| Atb2 | Ungwatiri | 36 | 16.9 | 1950 | 1981 | 2.66 | 9.82 |
| Atb3 | Abu-Quta | 32.7 | 14.88 | 1948 | 1987 | 2.06 | 17.76 |
| Atb4 | Haiya | 36.37 | 18.33 | 1950 | 1981 | 2.67 | 5.55 |
| Atb5 | Gedaref | 35.4 | 14.03 | 1903 | 1996 | 1.5 | 51.78 |
| Atb6 | Ghadambaliya | 34.98 | 14.2 | 1948 | 1988 | 1.65 | 42.88 |
| Blue Nile | | | | | | | |
| Blu1 | Bahr Dar | 37.41 | 11.6 | 1964 | 2004 | 1.3 | 120.9 |
| Blu2 | Debremarcos | 37.67 | 10.33 | 1964 | 2004 | 1 | 109.3 |
| Blu3 | Gonder | 37.4 | 12.55 | 1964 | 2004 | 1.21 | 97.43 |
| Blu4 | Addis Ababa | 38.75 | 9.03 | 1964 | 2004 | 1.02 | 99.91 |
| Blu5 | Kombolcha | 39.83 | 11.1 | 1964 | 2004 | 1.11 | 86 |
| Kyoga | | | | | | | |
| Kyo1 | Imanyiro | 33.27 | 0.29 | 1950 | 1977 | 0.65 | 96.33 |
| Kyo2 | Kapchorwa | 34.43 | 1.24 | 1950 | 1995 | 0.69 | 85.58 |
| Kyo3 | Buwabwale | 34.21 | 0.54 | 1950 | 1977 | 0.69 | 95.64 |
| Kyo4 | Ivukula | 33.35 | 0.57 | 1950 | 1997 | 0.68 | 105.7 |
| Long. : longitude (°) | | Lat. : latitude (°) | | LTM is in mm/month | | | |

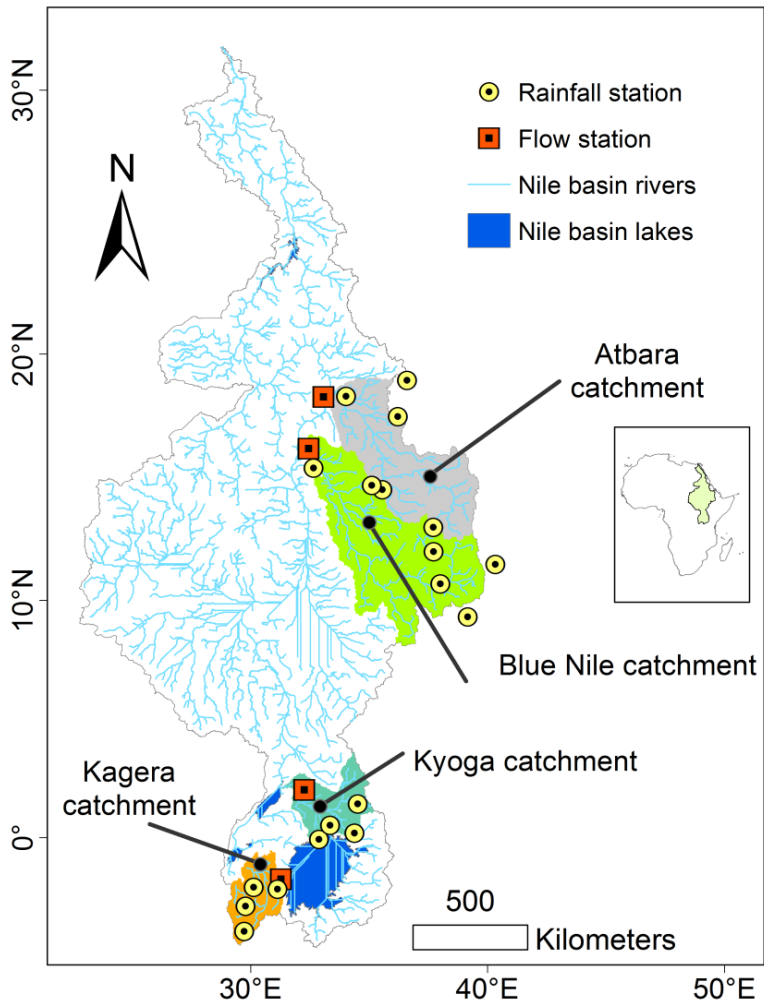


Figure 2.7 Selected catchments, and hydrological or meteorological stations (see Table 2.4 and Table 2.5 for details) in the Nile basin

2.2.2 River flow

a) Series for identification of the main driver of flow variation

River flow data recorded at the outlets of the four selected catchments (see Table 2.5 and Figure 2.7) in the Nile basin were obtained from various sources. The flow data at some stations were of both daily and monthly time scales. Monthly flow series for Stations 1, 3 and 4 were obtained from the Global Runoff Data Centre (GRDC), Koblenz, Germany. Monthly flow

series for Station 2 was obtained from the report by Kennedy & Donkin Power Ltd (1997). The monthly series were converted to annual time scale and the LTM (m³/s/year) and the CV of the annual flow are shown in the last two columns of Table 2.5. Whereas the monthly flow series were used for assessing their co-variation with rainfall, daily data were used for rainfall-runoff modeling.

Daily flow data for Station 1 of Table 2.5 was obtained at KU Leuven from the database of the FRIEND/Nile project. For Station 3, daily flow data were obtained based on personal connection. To check the similarity in variation of the data at the same station from the two sources, the daily series were converted to monthly time scale to conduct Pearson correlation test on the H_0 (no correlation). Based on the common periods of the monthly data from the two sources, *i.e.* 1965–1982 and 1950–1971 for the flow Khartoum and Kyaka Ferry, the coefficients of correlation were 0.98 and 0.93 which were greater than the critical values (at 0.1% level) of 0.15 and 0.20 respectively. Thus, the H_0 (no correlation between two datasets at the same station) was rejected at the 0.1% level for each of the stations.

Table 2.5 Selected hydrological stations and flow data

| SNo. | Name | Nile region | Country | Long. | Lat. | PMTS | PDTS | LTM | CV |
|--------------------------------|-------------|-------------------------------------|----------|-------|-------|---------------------|-----------|------|------|
| 1 | Kyaka Ferry | LVB | Tanzania | 31.42 | -1.27 | 1940-1971 | 1950-1986 | 194 | 0.35 |
| 2 | Kamdini | Kyoga Nile | Uganda | 32.33 | 2.24 | 1950-2000 | *** | 1075 | 0.31 |
| 3 | Khartoum | Blue Nile | Sudan | 32.51 | 15.64 | 1902-1982 | 1965-2002 | 1564 | 0.22 |
| 4 | Kilo 3 | Atbara | Sudan | 33.99 | 17.68 | 1912-1982 | *** | 385 | 0.35 |
| LVB : Lake Victoria basin | | Long. : Longitude (°) | | | | Lat. : Latitude (°) | | | |
| PMTS : period for monthly data | | PDTS : period for daily time series | | | | *** : No daily data | | | |

b) Series for analyses of flow quantiles and associated uncertainty

Daily discharge data from a total of 24 stations in the LVb were obtained from the FRIEND/Nile project. The locations of the data at these stations can be found presented in Figure 2.4 and Table 2.6 respectively. Long discharge time series, preferably above 25 years, were used. Table 2.6 shows, for these stations, the flow record length, coordinates, and area of the catchment upstream of each station. The flow metrics CV and long-term average computed based on the daily time scale are also shown in the last

two columns of the mentioned table. Although the CV for series at most stations was above 0.5, it notably ranged from 0.11 (Station 5) to 1.99 (Station 6). Each flow mean value for the selected data period fell within the range 1-270 m³/s and depended on the area of the catchment.

Table 2.6 Overview of the selected hydrological stations and flow data

| SNo | River catchment | Station ID | Area [km ²] | Data period | | Location | | CV | Mean [m ³ /s] |
|------------------------|-----------------|------------|-------------------------|-------------|------|----------|-------|------|--------------------------|
| | | | | From | To | Long. | Lat. | | |
| 1 | Biharamulo | *** | 1,981 | 1950 | 2004 | 31.29 | -2.62 | 1.25 | 18.2 |
| 2 | Bukora | 81270 | 8,392 | 1951 | 1976 | 31.48 | -0.85 | 1.41 | 3.1 |
| 3 | Grumeti | 5F3 | 13,363 | 1950 | 2004 | 33.94 | -2.06 | 0.43 | 11.2 |
| 4 | Gurcha-migori | 1KB05 | 6,600 | 1950 | 2004 | 34.2 | -0.95 | 0.67 | 57.7 |
| 5 | Isanga | 114012 | 6,812 | 1976 | 2004 | 32.77 | -3.21 | 0.11 | 29.7 |
| 6 | Kagera | 58370 | 54,260 | 1950 | 1994 | 31.43 | -1.29 | 1.99 | 266.3 |
| 7 | Katonga | 100006 | 15,244 | 1950 | 1975 | 31.95 | -0.09 | 0.65 | 5.1 |
| 8 | Koitobos | 1BE06 | 813 | 1949 | 1975 | 35.09 | 0.97 | 0.97 | 3.2 |
| 9 | Magogo-maome | 113012 | 5,207 | 1950 | 2004 | 33.15 | -2.92 | 0.28 | 7.8 |
| 10 | Mara | 107072 | 13,393 | 1950 | 2003 | 34.56 | -1.65 | 0.63 | 37.3 |
| 11 | Mbalangeti | 111012 | 3,591 | 1950 | 2004 | 33.86 | -2.22 | 0.50 | 4.3 |
| 12 | Moiben | 1BA01 | 188 | 1953 | 1990 | 35.44 | 0.8 | 0.50 | 1.3 |
| 13 | Nyakizumba | 100005 | 359 | 1950 | 1987 | 30.08 | -1.32 | 1.26 | 4.7 |
| 14 | Nyando | 1GD01 | 3,652 | 1962 | 2001 | 35.04 | -0.10 | 0.68 | 20.3 |
| 15 | Nyangores | 1LA03 | 4,683 | 1963 | 1993 | 35.35 | -0.79 | 1.18 | 8.3 |
| 16 | Nzoia | 1EF01 | 12,676 | 1974 | 1999 | 34.08 | 0.13 | 1.20 | 115.5 |
| 17 | Ogilla | 1GD03 | 2,650 | 1970 | 1996 | 34.96 | -0.13 | 0.59 | 16.5 |
| 18 | Ruizi | 100004 | 2,070 | 1970 | 1998 | 30.65 | -0.62 | 0.58 | 3.8 |
| 19 | Sergoit | 1CA02 | 659 | 1959 | 1990 | 35.06 | 0.63 | 0.52 | 2.3 |
| 20 | Simiyu Ndagalu | 5D1 | 1,205 | 1970 | 1996 | 33.56 | -2.63 | 0.63 | 11.8 |
| 21 | Sio | 1AH01 | 1,450 | 1958 | 2000 | 34.15 | 0.38 | 0.88 | 11.4 |
| 22 | Sondu | 1JG01 | 3,508 | 1950 | 1990 | 35.01 | -0.39 | 0.90 | 43.3 |
| 23 | South Awach | 1HE01 | 3,156 | 1950 | 2004 | 34.54 | -0.47 | 1.36 | 6.0 |
| 24 | Yala | 1FG01 | 3,351 | 1950 | 2000 | 34.51 | 0.09 | 1.43 | 37.7 |
| SNo: Station number | | | Long. : Longitude [°] | | | | | | |
| *** Missing station ID | | | Lat. : Latitude [°] | | | | | | |

2.2.3 Potential evapotranspiration

In investigating the main driver of river flow variation in the selected catchments of Kagera and Blue Nile, potential evapotranspiration (PET) was required for the rainfall-runoff modeling. Because long-term evaporation and temperature data are lacking from the study area, reanalysis series were deemed suitable to estimate the daily PET. For both

catchments (Kagera and Blue Nile), air temperature for PET estimation was required to cover the period 1950-2002.

Daily air temperature could be obtained from various existing reanalyses data. However, most of the reanalyses data start after 1950 *e.g.* ERA-40 (from 1957 to 2002), ERA-Interim (from 1979 to present), the Japanese 55-year Reanalysis JRA-55 (from 1958 to 2013), *etc.* Eventually, the PGF (Sheffield *et al.*, 2006) gridded ($0.5^\circ \times 0.5^\circ$) daily maximum (T_{\max}) and minimum (T_{\min}) temperature series covering the period 1948-2010 were obtained from the same data source as that for rainfall which was already given in Section 2.2.1c. The PET was computed using the FAO Penman-Monteith method (Allen *et al.*, 1998) considering T_{\min} and T_{\max} . For data scarce regions *e.g.* the study area, the procedure adopted for estimating the PET for daily and 10-day time scales amidst missing data can be found in the FAO Irrigation and Drainage Paper 56.

2.2.4 Series related to large-scale ocean-atmosphere interactions

Increase and decrease in the atmospheric pressure at sea level *i.e.* Sea Level Pressure (SLP) can reveal useful information on atmospheric circulation which brings about dry and wet conditions, respectively. Changes in Sea Surface Temperature (SST), on the other hand, can generate imbalance in the heat-flux field which can also bring about anomalous atmospheric circulation and rainfall patterns (Horel, 1982). To gain an insight into the consequence, on rainfall variability, of pressure changes (occurring over the different oceans) and circulation due to SST anomalies, a number of readily and freely available series (Table 2.7) related to large-scale ocean-atmosphere interactions were used.

a) SLP and related series

1. The global monthly historical SLP (HadSLP2) data (Allan and Ansell, 2006) from reconstruction of land and marine observations were obtained from the Physical Sciences Division (PSD) of the National Oceanic & Atmospheric Administration (NOAA)/Earth System Research Laboratory (ESRL) (see Table 2.7).

2. The North Atlantic Oscillation (NAO) index is the normalized SLP difference between SW Iceland (Reykjavik), Gibraltar and Ponta Delgada (Azores). The NAO index (Hurrell, 1995; Jones *et al.*, 1997) was obtained from the Climatic Research Unit (CRU).
3. The North Pacific Index (NPI) is the area-weighted SLP over the region 30–65° N, 160° E–140° W. The NPI (Trenberth and Hurrell, 1994) was obtained from the National Centre for Atmospheric Research (NCAR)/University Corporation for Atmospheric Research (UCAR).
4. The Trans Polar Index (TPI) is defined as the normalized pressure difference between Hobart (Australia) and Stanley (Falkland Islands). The TPI (Jones *et al.*, 1999; Pittock, 1980; 1984) was obtained from the CRU.
5. The Southern Oscillation Index (SOI) is defined as the normalized pressure difference between Tahiti and Darwin, Australia. The SOI based on Ropelewski and Jones (1987) was obtained from the CRU.

b) SST and related series

1. The global monthly historical SST (HadSST2) anomalies with respect to the 1961–1990 climatology (Rayner *et al.*, 2006) were obtained from the UK Met Office through the British Atmospheric Data Centre (BADC).
2. The Atlantic Multi-decadal Oscillation (AMO) index is defined as the SST averaged over 25–60° N, 7–70° W minus the regression on global mean temperature (van Oldenborgh *et al.*, 2009). The AMO index was obtained from the Koninklijk Nederlands Meteorologisch Instituut (KNMI) through their climate explorer.
3. The Pacific Decadal Oscillation (PDO) index is the leading principal component of North Pacific monthly SST variability (poleward of 20° N in the Pacific basin). The digital values of the PDO index of Mantua *et al.* (1997) were obtained from the database of the Joint Institute for the Study of the Atmosphere and Ocean.
4. The Indian Ocean Dipole (IOD) index is the anomalous SST difference between the Western (50° to 70° E and 10° S to 10° N) and the South

Eastern (90° to 110° E and 10° S to 0° N) Equatorial Indian Ocean. The IOD index was obtained from the Japan Agency for Marine-Earth Science and Technology.

5. Area averaged Niño SST indices (Rayner *et al.*, 2003; Trenberth, 1997) for the Tropical Pacific regions described by Niño 3 (90 to 150° W and 5° N to 5° S), Niño 3.4 (120 to 170° W and 5° N to 5° S), and Niño 4 (150° W to 160° E and 5° N to 5° S) were obtained from the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration.

Table 2.7 Overview of SLP, SST and related time series

| SNo | Data | Downloaded from: | Accessed on | Period |
|---------------------------|----------|---|-------------|-----------|
| SLP related series | | | | |
| 1 | HadSLP2 | http://www.esrl.noaa.gov/psd/ | 12/5/2014 | 1850-2013 |
| 2 | NAO | http://www.cru.uea.ac.uk/cru/data/nao/ | 29/01/2013 | 1900-2000 |
| 3 | NPI | http://www.cgd.ucar.edu/cas/jhurrell/indices.data.html | 29/01/2013 | 1900-2011 |
| 4 | TPI | http://www.cru.uea.ac.uk/cru/data/tpi/ | 29/01/2013 | 1900-2003 |
| 5 | SOI | http://www.cru.uea.ac.uk/cru/data/soi/ | 29/01/2013 | 1900-2004 |
| SST related series | | | | |
| 6 | HadSST2 | http://badc.nerc.ac.uk/ | 30/05/2014 | 1850-2013 |
| 7 | AMO | http://climexp.knmi.nl/data/iamo_hadsst2.dat | 29/01/2013 | 1900-2012 |
| 8 | PDO | http://jisao.washington.edu/pdo/PDO.latest | 30/01/2013 | 1900-2011 |
| 9 | IOD | http://www.jamstec.go.jp/frcgc/research/d1/iod | 20/01/2014 | 1900-2003 |
| 10 | Niño 3 | http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino3/ | 29/01/2013 | 1900-2011 |
| 11 | Niño 3.4 | http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/ | 29/01/2013 | 1900-2011 |
| 12 | Niño 4 | http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino4/ | 29/01/2013 | 1900-2012 |
| SNo: Serial Number | | | | |

Chapter Three

Variability in seasonal and annual rainfall

This chapter is based on the paper:

Onyutha C, Willems P (2015) Spatial and temporal variability of rainfall in the Nile basin. Hydrology and Earth System Sciences, 19(5):2227-2246

3.1 Introduction

Whilst the warming and changes in all components of the climate system are hypothetically ascribed to the continued increase in greenhouse gas emissions (IPCC, 2013), annual rainfall variability may be driven by large-scale ocean-atmosphere interactions (Fowler and Archer, 2005). Rainfall variability can also be influenced by regional features such as water bodies, topography, transition in land cover and/or use, *etc.* According to Gleick and Adams (2000), climate variability has become a major concern for water resources managers and decision makers. Climate variability appears to have a very marked effect on many hydrological series (Kundzewicz and Robson, 2004). Among all the atmospheric variables, variability in rainfall is a critical factor in determining the spatio-temporal influence of the climate system on hydrology and related water management applications. It is vital to study the historical temporal variability in rainfall, and associated possible driving forces which might be from large-scale ocean-atmosphere interactions, anthropogenic factors as well as the influence from regional features such as water bodies, topography, *etc.*

Different methods exist for computing rainfall variability; some being the Empirical Orthogonal Functions (EOF) and Autocorrelation Spectral Analysis (ASA). The EOF-based method applies Principal Component (PC) analysis to a group of rainfall time series data to extract coherent variations that are dominant. It entails computation of eigenvectors and eigenvalues of a covariance or correlation matrix obtained from a group of original rainfall time series data. Some of the studies on rainfall of the Nile basin that used the method of EOF include Indeje *et al.* (2000), Ogallo (1989), and

Semazzi and Indeje (1999). The ASA (Blackman and Tukey, 1959; World Meteorological Organization, 1966) utilizes the connection of autocovariance estimation and spectral analysis by the Fourier transform. In the study area, this method was applied by Nicholson and Entekhabi (1986) to assess the quasi-periodic behavior of rainfall variability in Africa and its relationship to Southern Oscillation. In this paper, Quantile Perturbation Method (QPM) was adopted. The QPM is a quantile anomaly indicator which interestingly combines the frequency of extreme events and perturbation to rainfall extremes as applied by Ntegeka and Willems (2008) to rainfall extremes at Uccle in Belgium. This method was selected because of its robustness to quantify variability in hydro-climatic variables. In the study area, the usefulness of the QPM has been demonstrated by Mbungu *et al.* (2012), Nyeko-Ogiramoi *et al.* (2013), Onyutha and Willems (2015a; 2015b), and Taye and Willems (2011, 2012) but limited to specific catchments in the Nile basin. Outside the study area, this method was also applied by Willems (2013) for rainfall extremes in Europe, and Mora and Willems (2012) for the Andes region of Ecuador. The shortcomings of the above recent studies that applied the QPM on the rainfall variability of the Nile basin as addressed in this study include: 1) lack of attempts to investigate any possible linkages of the rainfall variability to large-scale ocean-atmosphere interactions by Mbungu *et al.* (2012), and Onyutha and Willems (2015a; 2015b), (2) the use of few climate indices or series to explain rainfall variability by Nyeko-Ogiramoi *et al.* (2013), and Taye and Willems (2012), and (3) the limitation of the variability study to sub catchments of the Nile basin. In line with the shortcoming (3), considering the entire River Nile basin would be helpful in understanding the regional differences in the rainfall statistics. This is vital in regional planning for the management of agricultural practices given that subsistence and rain-fed agriculture, together with variability EHRI is one of the main causes of food insecurity and the most daunting challenge the entire Nile basin faces (Melesse *et al.*, 2011).

Therefore, this study was aimed at analyzing spatio-temporal variability of rainfall over the Nile basin while investigating any linkages to large-scale ocean-atmosphere interactions. This was done at both seasonal and annual

scales. Variability of rainfall volumes at such time scales was deemed important for the estimation of crop water requirements for agricultural practices.

3.2 Methodology

3.2.1 Computing variability using the QPM

The QPM considers changes in quantiles with similar exceedance frequency or return period T compared from two time series. One of the series is taken to be the complete time series, while the other is a time slice of the entire historical period. To apply QPM to an annual series (say, variable y) of record length equal to N' years, the following steps can be followed:

- i) Select a block size (time slice) of length D (in years) and also choose the threshold exceedance probability T_t .
- ii) Put the window of length D at the beginning of the series y .
- iii) Denote the subseries covered by the window in Step (ii) as x .
- iv) Rank x in descending order such that from the highest to the lowest value, the rank j will be given by $j = 1, 2, \dots, D$.
- v) Rank the full time series y also in descending order to obtain $j = 1, 2, \dots, N'$ as the rank from the highest to the lowest value.
- vi) Compute the exceedance probability or the empirical return period T_x for the ranked series x in Step (iv) to obtain (D/j) i.e. $D/1, D/2, \dots, D/D$.
- vii) Compute the empirical return period T_y for the ranked series y in Step (v) to obtain (N'/j) i.e. $N'/1, N'/2, \dots, N'/N'$.
- viii) Match values x and y which have similar exceedance probabilities. For instance, considering the highest ranking value of x i.e. with $T_x = D/1$, the corresponding value of y would be the one from Step (v) with $T_y = D/1$. If D and N' are 10 and 50 years respectively, the highest ranking value of x would be compared with the fifth highest value of y because $T_x = 10/1$ is equal to $T_y = 50/5$. For x with T_x which does not match the exact value of T_v , interpolation is done to locate y for which $T_y = T_x$.
- ix) Compute the relative change as the ratio of x to y with the same exceedance probability. The ultimate anomaly value for the time slice

period of Step (ii) is obtained as the average value of the perturbation factors above the threshold selected from Step (i).

- x) Determine the time of observation which corresponds to the mid-point of the window covering the subseries x and call it $M^\#$.
- xi) Slide the window in Step (ii) by one time step and repeat Steps (iii) to (x). The time step for annual series is taken as one year.
- xii) For the complete analysis accounting for all possible times obtained from the repeated process of moving slice over the entire time series, the temporal anomalies are plotted against corresponding values of $M^\#$ from Step (x).

Due to sensitivity of the QPM to block length (*i.e.* a subseries of the full time series covering the period of interest), preliminary analysis was conducted using block periods of 5, 10 and 15 years. A block length of 15 years was found to give a much clearer oscillation pattern in the rainfall data than for 5 and 10 years, and was eventually adopted for this study. Alongside details on the QPM, the selection and sensitivity of the block length can be found fully discussed by Ntegeka and Willems (2008), and/or Willems (2013).

For the variability analyses, the QPM was applied to series of:

1. annual rainfall totals and mean of SLP, SST or related variables in each year, and
2. seasonal rainfall totals and mean of SLP, SST or related variables in each season.

3.2.2 Test of significance of the QPM anomalies

The H_0 that the observed temporal variability in the extreme rainfall events occur in a random way *i.e.* by only natural variability (*i.e.* there is no persistence in the temporal climate variation) was considered. To verify H_0 at level of significance α_s of 5%, nonparametric bootstrapping Monte Carlo simulations were used to test the statistical significance of the temporal variation. To derive bounds of variability using 95% Confidence Interval (CI), nonparametric bootstrapping method (Davidson and Hinkley, 1997) was employed as follows:

1. The original full time series was randomly shuffled to obtain a new temporal sequence.
2. The new series was divided to obtain subseries each of length equal to the time slice period.
3. New temporal variation of anomalies was obtained by applying QPM to the shuffled series.
4. Steps 1, 2 and 3 were repeated 1000 times to obtain 1000 anomaly factors for each subseries.
5. The anomaly values were ranked from the highest to the lowest.
6. The limits (upper, lower) of the 95% CI for each time moment were taken as the (25th, 975th) anomaly values.

Ideally, to preserve the dependency in time of the series, the shuffling in Step (1) would have been done in blocks. However, this study focused on extreme values thereby meaning that not all the shuffled values were used to compute anomalies. Thus, the QPM results were not greatly affected by the random shuffling.

To interpret the results of QPM, the average of the long-term or full time series is taken as the reference (*i.e.* anomaly of zero). Values above and below this reference (also referred to as oscillation high (OH) and low (OL) respectively) are considered to characterize the variability in the given series. An anomaly of, say, 5% (or -5%) for a given year indicates that the extreme quantiles in the sub-period centered on this year was on average 5% higher (or lower) than the reference. If the upper (or lower) 95% CI limit is up crossed (or down crossed) by the anomaly values in the OH (or OL) period, it means that the OH (or OL) is statistically significant in the rainfall of the selected station.

3.2.3 Spatial differences in rainfall anomalies

The temporal patterns of anomaly (QPM results) in the rainfall of the different selected stations were compared. This was also graphically done by examining the similarities in the co-occurrences of the OHs and OLs. Strong spatial differences of these rainfall statistics could indicate difference in the possible driving forces of temporal variability across the

study area. It could also help to indicate the periods during which a particular region was characterized by dry or wet spells.

3.2.4 Correlation analysis

Any possible linkage of the rainfall variability to large-scale ocean-atmosphere interactions was sought using Pearson correlation test (at the significance levels of 5 and 1%) under the H_0 "there is no correlation between the rainfall QPM results and those of the SLP, SST or climate indices". To locate the part of the world over which the driving influence for temporal variability in rainfall over the Nile basin originates, correlation between the regional or basin-wide annual/seasonal total rainfall QPM results and corresponding SLP or SST at each grid point was calculated. Next to the correlation with the SLP, also the correlation with SLP differences, being a measure of atmospheric circulation, was tested. For that purpose, the SLP difference were considered over any two regions with significant correlations but of opposite signs and for which it is expected that the related atmospheric circulation drives the rainfall over the region under study. The oscillation pattern of the SLP difference was compared with that of the regional/basin-wide rainfall total. To further confirm the results of the comparison, correlation analyses between the QPM results of station-based annual/seasonal rainfall and those of climate indices were analyzed. Coefficients of correlation between rainfall QPM results and those of the climate indices were analyzed for further investigation.

To investigate the influence of data record length on the rainfall variability analysis, correlation was done in two ways. First, the analysis was done using short-term series selected over the period when all the stations had data. In the second procedure, correlation analysis was done using the full time series as presented in Table 2.1.

3.3 Results and discussion

3.3.1 Differences in temporal anomalies of rainfall

Figure 3.1 shows temporal variability of quantile anomaly (QPM results) for annual rainfall at the different stations of each group. Just like for the long-term mean monthly rainfall (Figure 2.3), it is shown that these general patterns of the temporal variability at stations of each group are similar (Figure 3.1).

For group A, the OL occurred from the late 1940s to 1950s. The OL of this period was not significant at the 5% level in any of the selected stations. The period from 1960 to mid-1980s was characterized by its change in annual rainfall above the reference; this OH was significant at Stations 2, 4 and 7 of Table 2.1. This is consistent with the findings of Kizza *et al.* (2009) that for the LVB, there was a significant step jump in mean of annual rainfall in the 1960s. Generally, the oscillation patterns of the annual rainfall QPM results at stations of group A are consistent with the findings of Mbungu *et al.* (2012) and Nyeko-Ogiramoi *et al.* (2013) for the LVB. For group B, generally the period 1920s to 1960 was characterized by anomalies above reference. This OH was significant at Stations 9–11, 14–15, 19–20, 24, 26, 28, 30, and 32 of Table 2.1. The period from 1960s to 1980s was below reference and this OL was significant at Stations 10–13, 15, 17–21, 26–27, and 31–32 of Table 2.1. In line with this OL of the period 1960s to 1980s, Hulme (1992) found that the mean annual rainfall in the Sahel region for the decades 1970s and 1980s declined by 30%. For group C, the period of OL occurred in the late 1920s to 1930s. This was, however, based on only Station 36 whose record started in 1904. The OH occurred in the period from about 1940 to the early 1950s. This was significant at Stations 34–35 and 37 of Table 2.1. Though not significant at any station, OL occurred over the period of mid-1950s to 1970s. For the QPM results of rainfall in the main wet season (not shown) of each group, the stations of Table 2.1 at which there were significant OHs like for the annual rainfall (over the same period) were 2, 4 and 7 (MAM); 22–23, 26–27, and 32 (JJAS); 35–37 (ONDJF). However, there were some stations that exhibited significant OHs in the main wet season but not for the annual rainfall. From

Table 2.1, these included Stations 3 and 8 (for MAM rainfall of group A), 13 and 17 (for JJAS rainfall of group B). The OLs of the main seasons were insignificant at most stations except 18 and 20 over the periods 1922–1933 and 1920–1922 respectively.

Whereas one might think of the impacts of climate change on rainfall intensities, it is vital to note that the results in this study (Figure 3.1) showing decrease or increase in rainfall quantiles depict the decadal oscillations (based on short-term data) which may not be indicative of the long-term trends in the data as required for assessment of climate change.

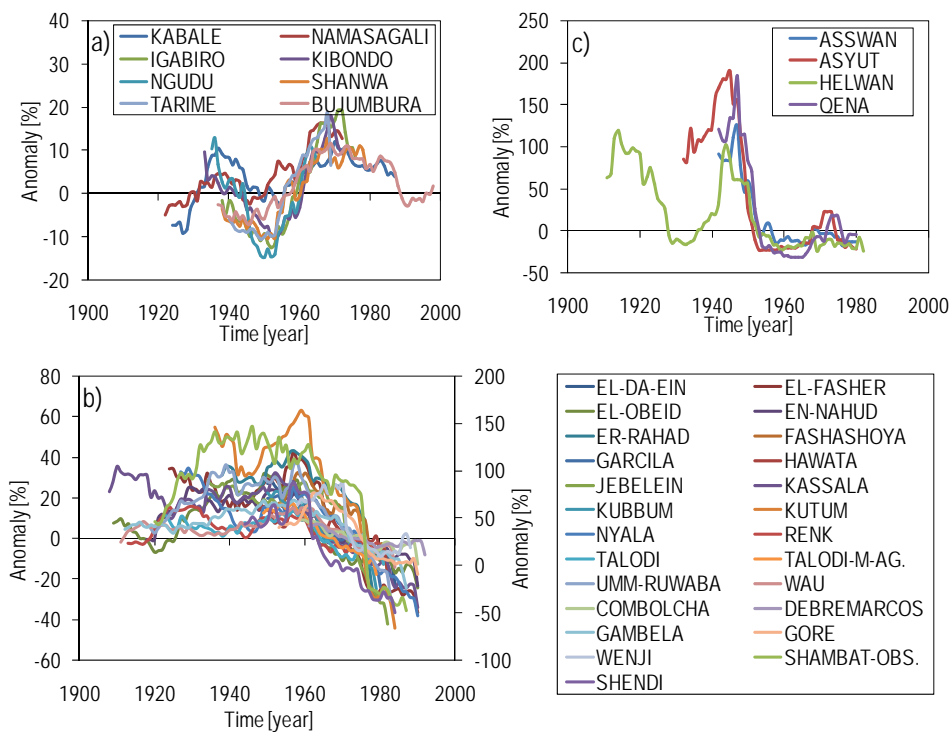


Figure 3.1 QPM results for annual rainfall using a time slice of 15 years and rainfall stations for group (a) A, (b) B, (c) C

3.3.2 Co-variation of the SLP and rainfall variability

Figure 3.2 shows locations over which possible influences for temporal variability of rainfall quantiles (QPM results) in the Nile basin originate as determined using HadSPL2 data. For the spatial coherence of the driving forces of rainfall variability, correlations for groups A to C are obtained over the periods in which each station had data records *i.e.* 1935–1970, 1954–1992 and 1945–1985, respectively. Eventually, the critical values of the correlation coefficients at the level of significance α_s of 5% (1%) for groups A to C were 0.33 (0.41), 0.32 (0.41), and 0.31 (0.41), respectively.

For group A, it is shown in Figure 3.2a that the quantile anomalies of the region-wide annual rainfall is correlated to that of the annual SLP in the Indian Ocean and North America (positively), and Pacific Ocean, and Southern Ocean (around Antarctica) (negatively). For group B, it is shown that the annual rainfall variability is positively linked to SLP at the Pacific Ocean, and negatively to that at the Indian Ocean. This correlation is consistent with the findings of Taye and Willems (2011) for the annual rainfall over the Ethiopian Highlands. According to Camberlin (1997) the part of the Nile basin that receives the summer rains, *i.e.* JJAS season, has strong connection with the Indian monsoon. A deepened monsoon low increases the pressure gradient with the South Atlantic Ocean, resulting into strengthened moist South-Westerlies/Westerlies over Ethiopia, Sudan and Uganda (Camberlin, 2009). Tropical Easterly jet associated with the intense Indian monsoon is another factor, especially for the rainfall over the Sahelian belt (Grist and Nicholson, 2001). For group C, negative correlations can be seen in around the Southern part of Latin America (*i.e.* Southern Atlantic, and Pacific Ocean); however, positive correlations are obtained with influence from the Southern Ocean (near Antarctica), and though over a small area, also in the Central Pacific Ocean.

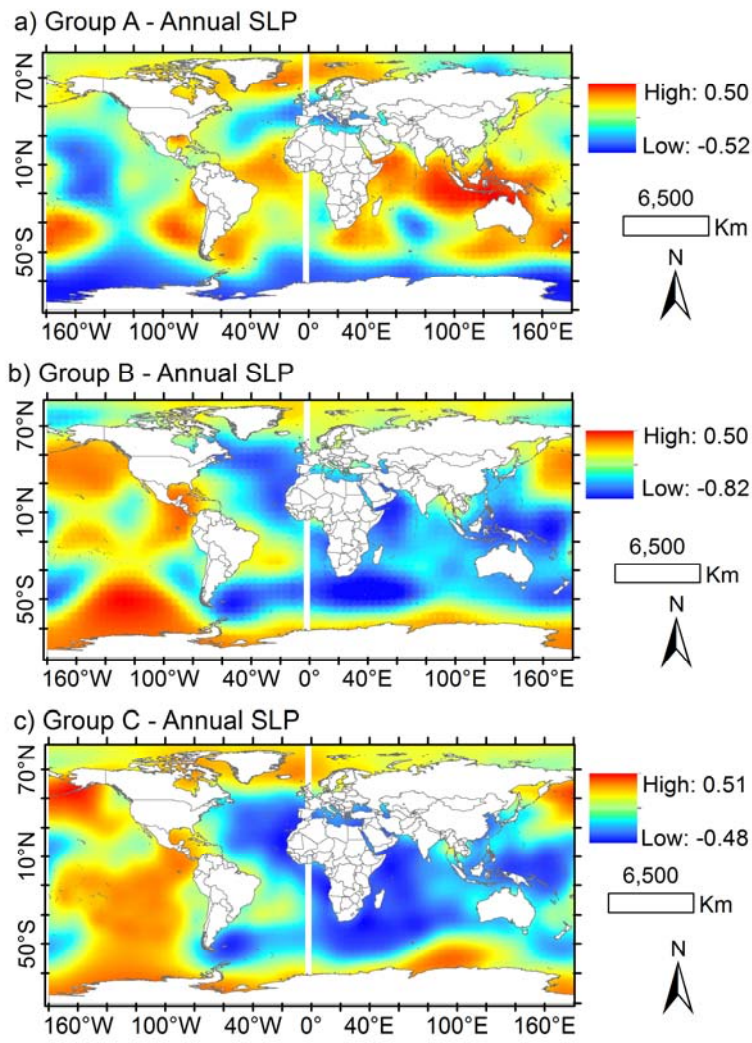


Figure 3.2 Correlation between anomaly in annual Sea Level Pressure (SLP) and that in region-wide annual rainfall for group (a) A, (b) B, (c) C

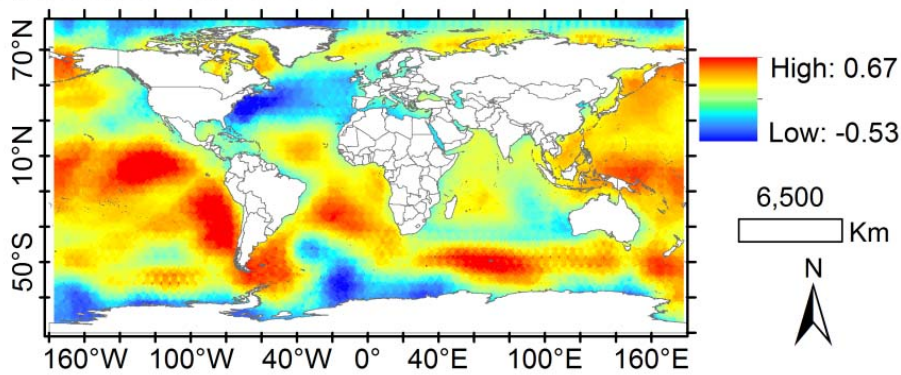
Figure 3.3 shows oceanic regions with the variability of their HadSST2 series correlated to that of the annual rainfall over the Nile basin. It is shown that positive influences for annual rainfall variability in the Equatorial region originate from the Pacific, Atlantic and Indian Oceans. For stations in groups B and C, correlations in similar oceanic locations are noted but with opposite sign to those in group A. For rainfall stations in Sudan and Ethiopia, the correlation signs over North Atlantic and South

Pacific are opposite. This is because rainfall variability of the Northern and Southern zones of Ethiopia are enhanced by an AMO and PDO SST pattern in warm and cool phase respectively (Jury, 2010). For the Equatorial region, it is shown that positive driving forces originate from the Pacific and Southern Indian Oceans. These locations are consistent to those indicated by Lyon and DeWitt (2012) and Williams and Funk (2011) from where influences on the rainfall variability over the Equatorial region originate. However, Tierney *et al.* (2013) showed that it is mainly the Indian Ocean SST which drives East African rainfall variability by altering the local Walker circulation, whereas the influence of the Pacific Ocean is minimal. For Sudan, Ethiopia and Egypt, it is shown that the SSTs from Southern parts of Indian and Pacific Oceans exhibit negative influence on the rainfall variability. These areas with negative correlations in Figure 3.3b and c lie from 30° to 60° S, and above 30° S respectively. The negative correlation between Indian Ocean SST and annual rainfall over Sudan and Ethiopia as seen in Figure 3.3b was also found in a number studies some of which include Korecha and Barnston (2007), Osman and Shamseldin (2002), and Seleshi and Demarée (1995).

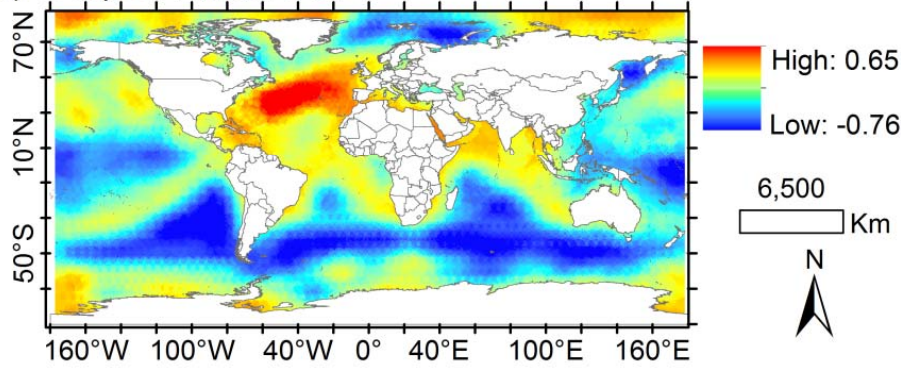
One may be interested in how the Indian Ocean influences rainfall over the study area. The SSTs of the Indian Ocean is different between the Eastern and the Western part. The difference in the SST (*i.e.* the IOD) tends to alter the atmospheric circulation thereby making the Western Indian Ocean to alternately get warmer and colder than the Eastern part. There are two phases of the IOD. In the positive phase (which is responsible for the rainfall over the study area), the Western part of Indian Ocean near East Africa is warmer than the normal condition. This warm ocean induces rising of the hot air (increased convection) leading to formation of clouds and heavy rainfall over the Western Indian Ocean and the surrounding area especially East Africa (*i.e.* the study area). In this positive IOD phase, the Eastern part is characterized by cooler than normal ocean temperature and thus, reduced convection. This creates less chance of rainfall over the Eastern part of the Indian Ocean and surrounding area. The negative IOD phase is characterized by the reversal of the processes of the positive phase. This leads to dry (wet) conditions in the Western (Eastern) part of

the Indian Ocean and the surrounding areas *e.g.* Somalia. Normally, the dry (wet) periods are also known to be linked to La Niña (El Niño).

a) Group A - Annual SST



b) Group B - Annual SST



c) Group C - Annual SST

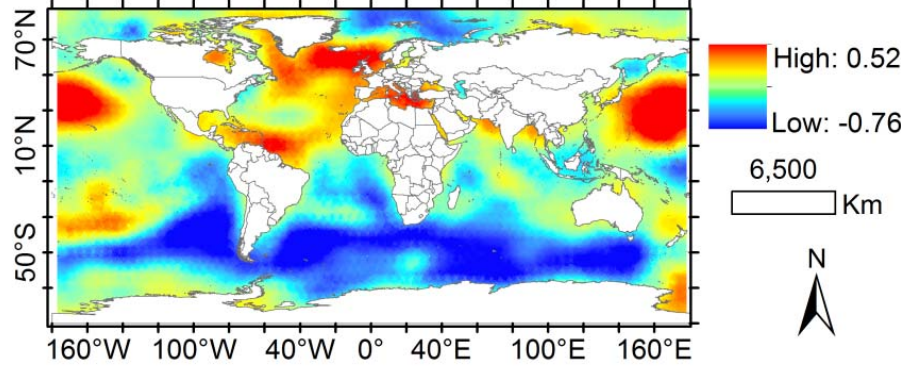


Figure 3.3 Correlation between annual Sea Surface Temperature (SST) and region-wide annual rainfall for group (a) A, (b) B, (c) C

3.3.3 Correlation analysis

Figure 3.4 shows temporal patterns of rainfall at selected stations of each group and SLP differences. The locations over which the SLP differences were taken were those with significant correlations but of opposite signs. According to Nicholson (1996), the wind and pressure patterns controlling the rainfall over the Equatorial region include a number of major air streams and convergence zones. The airstreams include the Northeast and Southeast monsoon which are relatively dry (unlike the SW monsoon of Asia), the contrastingly humid Congo air stream with Westerly and Southwesterly flow. The convergence zones include the ITCZ, the Congo Air boundary, and thirdly the zone that separates the dry, stable, Northerly flow of Saharan origin and the moist Southerly flow. In line with the above, the SLP differences were deemed to be taken at locations which fall on either side of the Nile basin *e.g.* between Indian and Atlantic Ocean, Indian and Pacific Ocean, *etc.*

For the same locations shown in Figure 3.4a–f, the values of the correlation between the SLP differences and rainfall for the different groups are shown in Table 3.1. The critical values of the correlation coefficients in Figure 3.4 and Table 3.1 are similar to those for Figure 3.2 and Figure 3.3. It can be seen that the patterns of the SLP differences capture the temporal oscillation in rainfall well. This may suggest that the decadal variability of rainfall over the Nile basin is caused by decadal variability in atmospheric circulation over basin that brings moist air from the different oceans *i.e.* Pacific and Indian Oceans, or Atlantic and Indian Oceans. However, the correlation (based on our short-term data) may merely indicate statistical measure of association. It may not, for various factors *e.g.* the high signal-to-noise ratio, be indicative of the actual dynamics of how the SLP and SST interact to bring about rainfall over the study area.

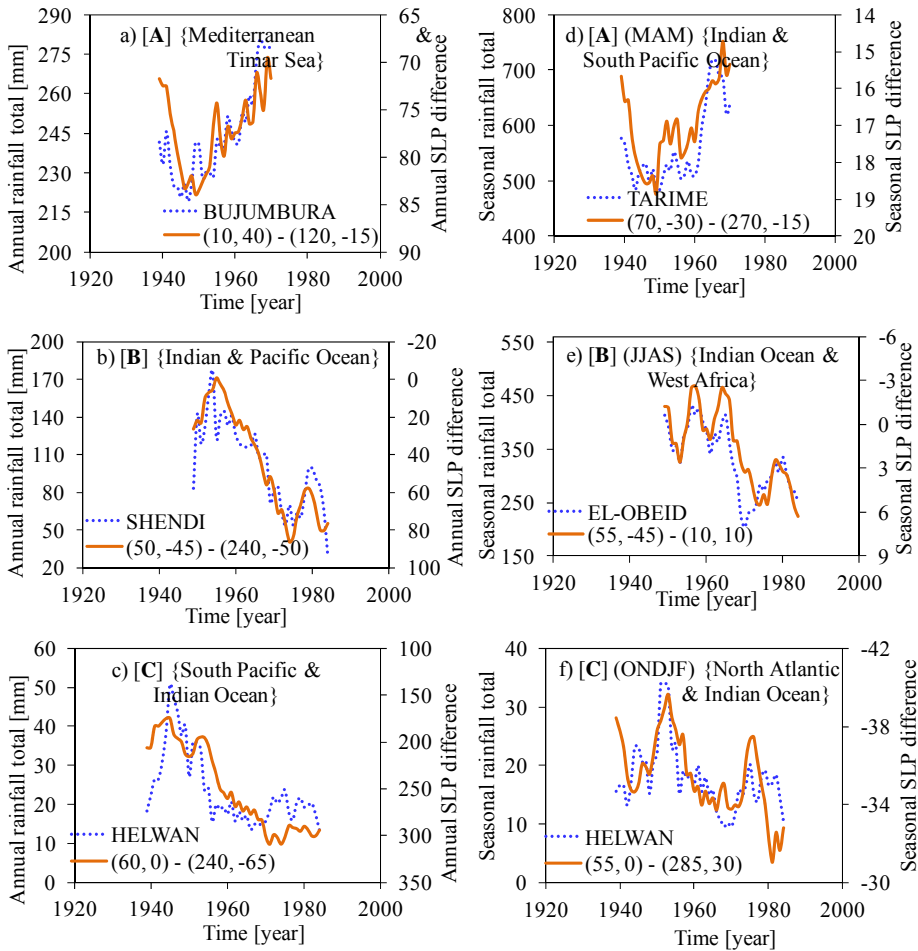


Figure 3.4 Annual SLP differences and rainfall at selected stations of the different groups A–C for a time slice of 5 years; the group labels are in []. The labels of the legend comprise the rainfall station and the coordinates (degree longitude and latitude) from where the SLP differences were taken. The label in "{ }" shows the locations where the coordinates are found. Annual and seasonal time scales are shown in charts (a)–(c) and (d)–(f) respectively.

Table 3.1 Correlation between rainfall and SLP differences taken at locations shown in Figure 3.4.

| Group-A | Annual | MAM | Group-B (contir | Annual | JJAS |
|------------|--------|-------|---------------------------|--------|-------|
| Kabale | -0.62 | 0.05 | Nyala | -0.69 | -0.71 |
| Namasagali | -0.4 | -0.6 | Renk | -0.63 | -0.42 |
| Igabi | -0.65 | -0.57 | Shambat-Obs. | -0.56 | -0.5 |
| Kibondo | -0.71 | -0.63 | Shendi | -0.88 | -0.75 |
| Ngudu | -0.53 | -0.5 | Talodi | -0.58 | -0.63 |
| Shanwa | -0.63 | -0.55 | Talodi-Min.Agri | -0.82 | -0.81 |
| Tarime | -0.5 | -0.81 | Umm-Ruwaba | -0.82 | -0.82 |
| Bujumbura | -0.68 | -0.61 | Wau | -0.76 | -0.83 |
| RWR | -0.75 | -0.85 | Combolcha | -0.71 | -0.64 |
| Group-B | Annual | JJAS | Debremarcos | -0.79 | -0.66 |
| El-Da-Ein | -0.86 | -0.78 | Gambela | -0.72 | -0.65 |
| El-Fasher | -0.87 | -0.63 | Gore | -0.5 | -0.32 |
| El-Obeid | -0.8 | -0.84 | Wenji | -0.14 | -0.13 |
| En-Nahud | -0.7 | -0.65 | RWR | -0.88 | -0.82 |
| Er-Rahad | -0.56 | -0.62 | Group-C | Annual | ONDJF |
| Fashashoya | -0.4 | -0.25 | Asswan | -0.53 | -0.2 |
| Garcila | -0.69 | -0.72 | Asyut | -0.66 | -0.2 |
| Hawata | -0.55 | -0.66 | Helwan | -0.72 | -0.59 |
| Jebelein | -0.64 | -0.55 | Qena | -0.59 | -0.28 |
| Kassala | -0.84 | -0.63 | RWR | -0.72 | -0.59 |
| Kubbum | -0.79 | -0.69 | | | |
| Kutum | -0.85 | -0.71 | RWR: Region-wide rainfall | | |

Figure 3.5 shows QPM results for a time slice of 15 years in the annual rainfall at Station 32 (from Table 2.1) and those for climate indices or time series over the period 1902-1996. It is shown for the selected station of group B that the primary influence is from the Atlantic Ocean as exhibited by the AMO index. This is confirmed by the correlation coefficients for all the group stations being positive and significant at significance levels of both 5 and 1%. According to Taye and Willems (2012), the changes in the Pacific and Atlantic Oceans are the major natural causes of the variability in the annual rainfall over the Blue Nile basin of Ethiopia. For the case of the AMO index, this is consistent, for the entire group B, with the finding of this study. However, for the influence from Pacific Ocean (in terms of PDO and SOI), it can be seen from Table 3.2 that the correlation coefficients for

Stations 11–12, 25 and 27 in Sudan are rather low. This could be because of the difference in regional influence from features such as topography, water bodies, transition in land cover and/or use, *etc.* According to Nicholson (1996), the highlands of the Rift Valley block the moist, unstable Westerly flow of the Congo air from reaching the coastal areas (especially of Ethiopia) leading to complex pattern of rainfall over the Ethiopian Highlands. For group B, it is also shown from Table 3.2 that influences also are obtained from Niño 3, the Indian Ocean and as far as the Extra-Tropical Southern Hemisphere. The annual rainfall in Sudan and Ethiopia (group B) has been shown in a number of studies to be influenced by the El-Niño Southern Oscillation (ENSO); see Osman and Shamseldin (2002) for Sudan; Fontaine and Janicot (1992), and Grist and Nicholson (2001) for the Sahel belt; Abtew *et al.* (2009), Beltrando and Camberlin (1993), Diro *et al.* (2011), Korecha and Barnston (2007), Seleshi and Demarée (1995), Segele and Lamb (2005), and Seleshi and Zanke (2004) for Ethiopia.

For group A, the primary influence appears to come from the North Atlantic Ocean as shown by the significant correlation (though negative) with the NAO index. This is consistent with the results of McHugh and Rogers (2001) who also found influence from the North Atlantic Ocean over the Southernmost part of the Nile basin (especially for rains in the December–February period). The sign of the correlation for NAO is also similar to that of AMO. Again, correlations are also obtained from Niño 3, Niño 3.4 and the Indian Ocean. Correlations between ENSO and rainfall for the Equatorial region have been even documented by Indeje *et al.* (2000), Nicholson (1996), Nicholson and Entekhabi (1986), Ogallo (1988), Nicholson and Kim (1997), Philipps and McIntyre (2000), Ropelewski and Halpert (1987), and Schreck and Semazzi (2004). Link between rainfall over the Equatorial region and the Indian Ocean was also found by Camberlin (1997) and Tierney *et al.* (2013). Nyeko-Ogiramoi *et al.* (2013) found that variability in the rainfall extremes in the LVB (within the Equatorial region) is also linked to the influence from the Pacific Ocean. However, as seen from Table 3.2, the influences from the Pacific Ocean are not very conclusive with respect to the correlations of NPI, PDO and SOI. The inconclusiveness of the correlations for group A and Pacific Ocean-related indices, instead agrees

with the remark of Tierney *et al.* (2013) as already mentioned for Figure 3.3.

For group C, it can be seen in Table 3.2 that the sign of the correlation for the each climate index is less consistent for all the stations than that obtained using SLP difference in Table 3.1. This suggests that conclusive correlation results with respect to the climate indices can perhaps be obtained using longer rainfall series. However, using Station 36 with the longest data record, positive significant influence is shown from the North Atlantic and the North Pacific Ocean.

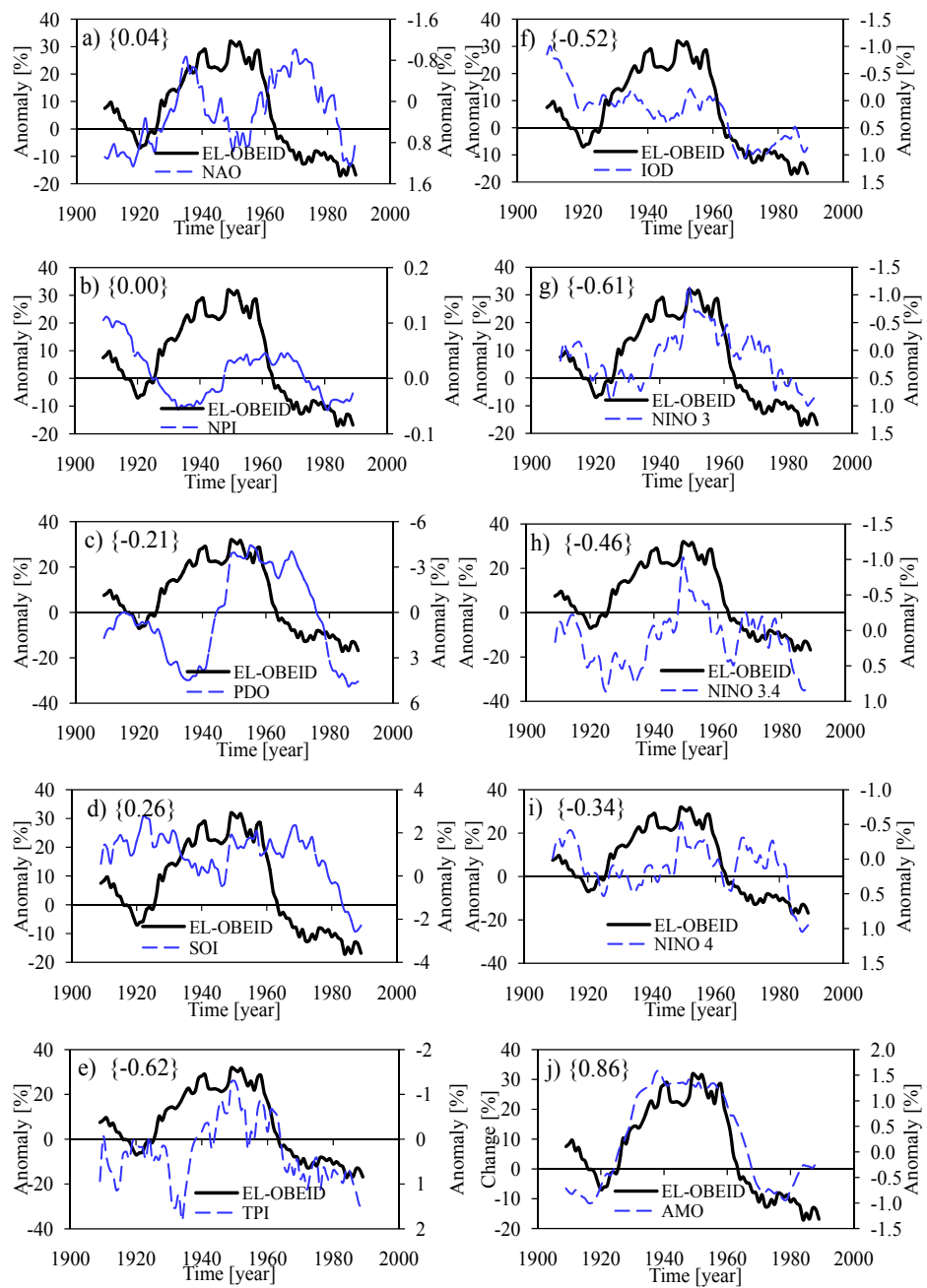


Figure 3.5 QPM results for annual rainfall and climate indices; the correlation coefficient between the two curves of each chart is put as label in "{ }".

Table 3.2 Correlation between QPM results for annual rainfall and climate indices over the periods1935–1970 (for group A), 1954–1992 (for group B) and 1945–1985 (for group C)

| SNo. | Station name | Climate indices and/or time series | | | | | | | | | Critical values | | |
|---|---------------|------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|----------------|----------------|
| | | NAO | NPI | PDO | SOI | TPI | IOD | Niño 3 | Niño 3.4 | Niño 4 | AMO | $\alpha_s=5\%$ | $\alpha_s=1\%$ |
| Group A | | | | | | | | | | | | | |
| 1 | Kabale | -0.70 | -0.13 | 0.07 | -0.33 | 0.29 | 0.54 | 0.04 | -0.04 | 0.03 | -0.23 | 0.25 | 0.33 |
| 2 | Namasagali | -0.64 | 0.48 | -0.49 | 0.08 | 0.15 | 0.60 | -0.22 | -0.12 | -0.18 | -0.32 | 0.28 | 0.37 |
| 3 | Igabi | -0.88 | 0.30 | -0.16 | 0.40 | 0.81 | 0.77 | 0.59 | 0.50 | 0.16 | -0.91 | 0.32 | 0.41 |
| 4 | Kibondo | -0.85 | 0.00 | 0.12 | 0.21 | 0.67 | 0.67 | 0.56 | 0.53 | 0.31 | -0.79 | 0.32 | 0.41 |
| 5 | Ngudu | -0.84 | -0.43 | 0.56 | -0.25 | 0.63 | -0.16 | 0.83 | 0.89 | 0.79 | -0.40 | 0.36 | 0.46 |
| 6 | Shanwa | -0.78 | 0.25 | -0.19 | 0.36 | 0.76 | 0.72 | 0.62 | 0.50 | 0.09 | -0.87 | 0.31 | 0.40 |
| 7 | Tarime | -0.81 | 0.56 | -0.38 | 0.43 | 0.58 | 0.42 | 0.49 | 0.60 | 0.28 | -0.90 | 0.37 | 0.47 |
| 8 | Bujumbura | -0.49 | 0.25 | -0.17 | 0.19 | 0.40 | 0.40 | 0.25 | 0.18 | -0.05 | -0.81 | 0.25 | 0.34 |
| Group B | | | | | | | | | | | | | |
| 9 | El-Da-Ein | 0.59 | 0.69 | -0.76 | 0.53 | -0.87 | -0.83 | -0.88 | -0.61 | -0.36 | 0.94 | 0.34 | 0.44 |
| 10 | El-Fasher | -0.01 | 0.42 | -0.41 | 0.67 | -0.56 | -0.80 | -0.51 | -0.23 | -0.33 | 0.75 | 0.24 | 0.32 |
| 11 | El-Obeid | 0.05 | 0.00 | -0.21 | 0.26 | -0.62 | -0.52 | -0.61 | -0.46 | -0.34 | 0.86 | 0.22 | 0.30 |
| 12 | En-Nahud | 0.02 | 0.07 | -0.20 | 0.51 | -0.53 | -0.79 | -0.48 | -0.30 | -0.34 | 0.84 | 0.23 | 0.31 |
| 13 | Er-Rahad | 0.44 | 0.26 | -0.22 | 0.08 | -0.62 | -0.78 | -0.59 | -0.22 | -0.07 | 0.76 | 0.31 | 0.41 |
| 14 | Fashashoya | 0.08 | 0.81 | -0.82 | 0.59 | -0.77 | -0.68 | -0.72 | -0.09 | 0.10 | 0.77 | 0.37 | 0.47 |
| 15 | Garcila | 0.62 | 0.69 | -0.81 | 0.42 | -0.81 | -0.80 | -0.88 | -0.51 | -0.17 | 0.91 | 0.36 | 0.46 |
| 16 | Hawata | 0.41 | 0.71 | -0.76 | 0.51 | -0.84 | -0.70 | -0.78 | -0.43 | -0.22 | 0.85 | 0.34 | 0.44 |
| 17 | Jebelein | 0.11 | 0.26 | -0.18 | 0.25 | -0.46 | -0.67 | -0.45 | 0.01 | 0.20 | 0.81 | 0.29 | 0.37 |
| 18 | Kassala | 0.27 | 0.48 | -0.24 | 0.56 | -0.45 | -0.80 | -0.51 | -0.39 | -0.53 | 0.39 | 0.22 | 0.30 |
| 19 | Kubbum | 0.76 | 0.61 | -0.75 | 0.23 | -0.86 | -0.90 | -0.83 | -0.45 | -0.14 | 0.96 | 0.37 | 0.47 |
| 20 | Kutum | 0.19 | 0.43 | -0.41 | 0.46 | -0.69 | -0.75 | -0.68 | -0.26 | -0.08 | 0.81 | 0.29 | 0.37 |
| 21 | Nyala | -0.27 | 0.32 | -0.33 | 0.81 | -0.35 | -0.70 | -0.43 | -0.21 | -0.43 | 0.62 | 0.25 | 0.33 |
| 22 | Renk | 0.35 | -0.28 | 0.19 | 0.27 | -0.27 | -0.53 | 0.03 | 0.18 | 0.26 | 0.57 | 0.24 | 0.32 |
| 23 | Shambat-Obs. | -0.17 | 0.24 | -0.34 | 0.45 | -0.47 | -0.53 | -0.58 | -0.31 | -0.28 | 0.77 | 0.24 | 0.32 |
| 24 | Shendi | 0.47 | 0.57 | -0.68 | 0.38 | -0.83 | -0.84 | -0.78 | -0.43 | -0.21 | 0.90 | 0.31 | 0.41 |
| 25 | Talodi | 0.67 | 0.10 | -0.06 | 0.14 | -0.56 | -0.86 | -0.33 | -0.09 | 0.05 | 0.76 | 0.26 | 0.34 |
| 26 | Talodi-M-Agr. | 0.76 | 0.65 | -0.79 | 0.28 | -0.83 | -0.90 | -0.81 | -0.45 | -0.22 | 0.98 | 0.36 | 0.46 |
| 27 | Umm-Ruwaba | 0.13 | 0.04 | -0.13 | 0.06 | -0.53 | -0.67 | -0.47 | -0.18 | 0.03 | 0.90 | 0.25 | 0.33 |
| 28 | Wau | 0.28 | 0.41 | -0.55 | 0.37 | -0.64 | -0.47 | -0.58 | -0.22 | -0.08 | 0.53 | 0.23 | 0.31 |
| 29 | Combolcha | -0.49 | 0.81 | -0.77 | 0.61 | -0.74 | -0.57 | -0.76 | -0.25 | -0.23 | 0.74 | 0.36 | 0.45 |
| 30 | Debre Marcos | -0.53 | 0.83 | -0.75 | 0.58 | -0.71 | -0.53 | -0.69 | -0.17 | -0.18 | 0.74 | 0.36 | 0.45 |
| 31 | Gambela | 0.00 | 0.32 | -0.52 | 0.46 | -0.65 | -0.46 | -0.74 | -0.45 | -0.37 | 0.71 | 0.23 | 0.31 |
| 32 | Gore | -0.62 | 0.92 | -0.89 | 0.80 | -0.56 | -0.32 | -0.75 | -0.41 | -0.43 | 0.44 | 0.33 | 0.42 |
| 33 | Wenji | -0.43 | 0.88 | -0.79 | 0.59 | -0.51 | -0.32 | -0.71 | -0.17 | -0.19 | 0.62 | 0.36 | 0.46 |
| Group C | | | | | | | | | | | | | |
| 34 | Asswan | 0.46 | -0.50 | 0.37 | -0.52 | -0.47 | -0.21 | -0.33 | -0.31 | 0.09 | 0.58 | 0.32 | 0.41 |
| 35 | Asyut | 0.09 | -0.71 | 0.66 | -0.59 | -0.05 | -0.23 | 0.13 | 0.16 | 0.39 | 0.52 | 0.29 | 0.37 |
| 36 | Helwan | 0.71 | 0.43 | 0.13 | 0.04 | -0.14 | -0.46 | -0.09 | -0.10 | -0.18 | -0.19 | 0.23 | 0.31 |
| 37 | Qena | 0.43 | -0.58 | 0.44 | -0.53 | -0.42 | -0.10 | -0.28 | -0.36 | 0.03 | 0.45 | 0.32 | 0.41 |
| Bold values indicate that H_0 (no correlation) was rejected at the significant at levels α_s of both 5% and 1% | | | | | | | | | | | | | |

Table 3.3 shows the linkages of seasonal rainfall to large-scale ocean-atmosphere interactions. Whereas the results in Table 3.3 are acceptably consistent with those from Table 3.2, it can be noted that the station to station difference in the magnitudes of the correlation become more

evident *e.g.* MAM rainfall of group A and NAO, IOD, Niño 3 and AMO. The fluctuations in the annual rainfall are more smoothened than in those of the seasonal rains. This could make the correlation signs and magnitudes clearer in annual than seasonal rainfalls. It could as well indicate that the driving forces of rains over some seasons are different from those of annual rainfall. In the Equatorial region, it is known that the main rains are less variable, so the inter-annual variability is related primarily to fluctuations in the short rains *i.e.* the October to December season (Nicholson, 1996). However, in this study the MAM (main) season was considered because of the direct relevance of its rainfall totals for agricultural management.

It is shown in Table 3.3 that the MAM rainfall of group A is linked to influence from Atlantic and Indian Oceans. The JJAS rainfall of group B is correlated with AMO, Niño 3, TPI and IOD. The ONDJF rainfall of group C is correlated with influences in the Indian Ocean.

Table 3.3 Correlation between QPM results for seasonal rainfall and climate indices considering full series record lengths as shown in Table 2.1

| SNo. | Station name | Climate indices and/or time series | | | | | | | | | Critical values | | |
|---|---------------|------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|----------------|----------------|
| | | NAO | NPI | PDO | SOI | TPI | IOD | Niño 3 | Niño 3.4 | Niño 4 | AMO | $\alpha_5=5\%$ | $\alpha_5=1\%$ |
| Group A (MAM season) | | | | | | | | | | | | | |
| 1 | Kabale | -0.40 | -0.40 | 0.27 | -0.43 | 0.14 | 0.39 | 0.00 | -0.18 | -0.02 | -0.05 | 0.25 | 0.33 |
| 2 | Namasagali | -0.57 | 0.46 | -0.36 | 0.39 | 0.43 | 0.58 | 0.16 | 0.24 | 0.06 | -0.71 | 0.28 | 0.37 |
| 3 | Igabiro | -0.83 | -0.04 | 0.13 | 0.15 | 0.79 | 0.89 | 0.62 | 0.43 | 0.19 | -0.86 | 0.32 | 0.41 |
| 4 | Kibondo | -0.60 | 0.31 | -0.24 | 0.32 | 0.42 | 0.81 | 0.20 | 0.18 | 0.03 | -0.92 | 0.32 | 0.41 |
| 5 | Ngudu | -0.73 | -0.69 | 0.77 | -0.34 | 0.78 | -0.25 | 0.85 | 0.79 | 0.69 | 0.06 | 0.36 | 0.46 |
| 6 | Shanwa | 0.25 | 0.46 | -0.56 | 0.30 | -0.12 | -0.20 | -0.19 | -0.06 | -0.12 | 0.04 | 0.31 | 0.40 |
| 7 | Tarime | -0.79 | 0.49 | -0.30 | 0.39 | 0.45 | 0.27 | 0.50 | 0.62 | 0.33 | -0.80 | 0.37 | 0.47 |
| 8 | Bujumbura | -0.14 | 0.03 | 0.09 | -0.18 | 0.46 | 0.36 | 0.46 | 0.45 | 0.30 | -0.73 | 0.25 | 0.34 |
| Group B (JJAS season) | | | | | | | | | | | | | |
| 9 | El-Da-Ein | 0.60 | 0.68 | -0.73 | 0.50 | -0.88 | -0.87 | -0.86 | -0.55 | -0.31 | 0.96 | 0.34 | 0.44 |
| 10 | El-Fasher | -0.02 | 0.43 | -0.41 | 0.69 | -0.53 | -0.80 | -0.49 | -0.21 | -0.33 | 0.73 | 0.24 | 0.32 |
| 11 | El-Obeid | 0.00 | 0.05 | -0.25 | 0.32 | -0.61 | -0.55 | -0.62 | -0.46 | -0.38 | 0.83 | 0.22 | 0.30 |
| 12 | En-Nahud | -0.12 | 0.27 | -0.45 | 0.62 | -0.60 | -0.69 | -0.64 | -0.45 | -0.50 | 0.77 | 0.23 | 0.31 |
| 13 | Er-Rahad | 0.15 | 0.53 | -0.43 | 0.28 | -0.45 | -0.59 | -0.45 | -0.02 | -0.02 | 0.50 | 0.31 | 0.41 |
| 14 | Fashashoya | 0.09 | 0.78 | -0.79 | 0.55 | -0.79 | -0.70 | -0.69 | -0.05 | 0.13 | 0.78 | 0.37 | 0.47 |
| 15 | Garcila | 0.64 | 0.67 | -0.79 | 0.39 | -0.81 | -0.82 | -0.87 | -0.50 | -0.15 | 0.92 | 0.36 | 0.46 |
| 16 | Hawata | 0.22 | 0.75 | -0.75 | 0.50 | -0.73 | -0.68 | -0.63 | -0.18 | -0.03 | 0.77 | 0.34 | 0.44 |
| 17 | Jebelein | 0.08 | 0.36 | -0.26 | 0.34 | -0.43 | -0.69 | -0.44 | 0.04 | 0.18 | 0.75 | 0.29 | 0.37 |
| 18 | Kassala | 0.22 | 0.48 | -0.23 | 0.64 | -0.39 | -0.78 | -0.46 | -0.33 | -0.53 | 0.33 | 0.22 | 0.30 |
| 19 | Kubbum | 0.74 | 0.63 | -0.74 | 0.20 | -0.87 | -0.93 | -0.80 | -0.38 | -0.08 | 0.97 | 0.37 | 0.47 |
| 20 | Kutum | 0.16 | 0.45 | -0.43 | 0.47 | -0.70 | -0.74 | -0.67 | -0.24 | -0.07 | 0.79 | 0.29 | 0.37 |
| 21 | Nyala | -0.25 | 0.43 | -0.43 | 0.82 | -0.42 | -0.70 | -0.50 | -0.26 | -0.46 | 0.61 | 0.25 | 0.33 |
| 22 | Renk | 0.32 | -0.25 | 0.06 | 0.26 | -0.35 | -0.44 | -0.09 | 0.04 | 0.16 | 0.56 | 0.24 | 0.32 |
| 23 | Shambat-Obs. | -0.49 | 0.13 | -0.10 | 0.49 | -0.09 | -0.42 | -0.17 | 0.11 | -0.04 | 0.46 | 0.24 | 0.32 |
| 24 | Shendi | 0.17 | 0.50 | -0.54 | 0.37 | -0.52 | -0.63 | -0.47 | -0.16 | -0.15 | 0.56 | 0.31 | 0.41 |
| 25 | Talodi | 0.62 | 0.25 | -0.19 | 0.19 | -0.56 | -0.81 | -0.32 | -0.05 | 0.07 | 0.67 | 0.26 | 0.34 |
| 26 | Talodi-M-Agr. | 0.67 | 0.70 | -0.79 | 0.24 | -0.78 | -0.90 | -0.71 | -0.29 | -0.09 | 0.97 | 0.36 | 0.46 |
| 27 | Umm-Ruwaba | 0.02 | 0.15 | -0.18 | 0.20 | -0.45 | -0.64 | -0.39 | -0.06 | 0.08 | 0.79 | 0.25 | 0.33 |
| 28 | Wau | 0.39 | 0.19 | -0.39 | 0.01 | -0.75 | -0.45 | -0.70 | -0.50 | -0.25 | 0.66 | 0.23 | 0.31 |
| 29 | Combolcha | -0.68 | 0.89 | -0.92 | 0.81 | -0.69 | -0.39 | -0.89 | -0.51 | -0.48 | 0.55 | 0.36 | 0.45 |
| 30 | Debremarcos | -0.33 | 0.68 | -0.57 | 0.41 | -0.75 | -0.71 | -0.57 | -0.04 | -0.03 | 0.86 | 0.36 | 0.45 |
| 31 | Gambela | 0.10 | 0.40 | -0.57 | 0.46 | -0.68 | -0.42 | -0.82 | -0.61 | -0.53 | 0.62 | 0.23 | 0.31 |
| 32 | Gore | -0.66 | 0.89 | -0.85 | 0.78 | -0.49 | -0.23 | -0.70 | -0.37 | -0.40 | 0.36 | 0.33 | 0.42 |
| 33 | Wenji | -0.41 | 0.87 | -0.78 | 0.59 | -0.45 | -0.26 | -0.70 | -0.20 | -0.22 | 0.58 | 0.36 | 0.46 |
| Group C (ONDJF season) | | | | | | | | | | | | | |
| 34 | Asswan | 0.39 | -0.17 | 0.16 | -0.38 | -0.47 | -0.54 | -0.24 | 0.05 | 0.25 | 0.66 | 0.32 | 0.41 |
| 35 | Asyut | -0.06 | -0.79 | 0.80 | -0.53 | 0.05 | -0.40 | 0.30 | 0.34 | 0.46 | 0.57 | 0.29 | 0.37 |
| 36 | Helwan | 0.80 | 0.55 | -0.01 | 0.28 | -0.09 | -0.51 | -0.01 | -0.01 | -0.21 | -0.28 | 0.23 | 0.31 |
| 37 | Qena | 0.56 | -0.48 | 0.36 | -0.53 | -0.56 | -0.30 | -0.39 | -0.37 | 0.05 | 0.63 | 0.32 | 0.41 |
| Bold values indicate that H_0 (no correlation) was rejected at the significance levels α_5 of both 5% and 1% | | | | | | | | | | | | | |

3.4 Conclusions and relevance of the findings

3.4.1 Answers to research questions

Is there variability in rainfall? Is the variability of a particular pattern? Is the variability significant? Is the variability linked to large-scale ocean-atmosphere interactions?

Rainfall series of seasonal and annual time scales were characterized by decadal (or multi-decadal) oscillations. The H_0 (natural randomness) was rejected the level of significance α_s of 5% in the rainfall of 17 out of 37 meteorological stations. On further analysis based on regional scale, the variability in rainfall was found to resonate well with the climate indices, SST or SLP from the various oceanic regions.

3.4.2 Relevance of the findings

These findings are vital for management of both water resources and agricultural practices.

The relevant information for the water managers is that rainfall wet and dry conditions can occur in clusters; which may require more considerate designs than if wet and dry conditions would occur fully randomly on a year-to-year basis. The frequency of hydro-climatic extremes during periods of dry and wet conditions can be taken into perspective. Besides, rainfall (design, *e.g.* intensity-duration-frequency relationship) statistics based on short periods may be biased. The possible rainfall variability drivers can be applied to bias-correct the statistics.

Rainfall variability drivers can also be used to predict an upcoming epoch of dry or wet conditions. During period of low rainfall totals, drought-tolerant crop varieties can be planted. Through sensitization, enough food can be stored from the cropping season preceding the predicted (upcoming) dry conditions. For rainfall totals above the normal condition, crop varieties with increased vigor but short developmental stages can be planted to expedite and increase yield. To help farmers select the right time for crop planting, further research should be conducted to investigate how rainfall variability affects crop production through seasonality shifts. This could be done by investigating whether rain starts early or late during the wet season of an epoch characterized by low or high rainfall total.

Chapter Four

Variability in extreme rainfall

This chapter is based on the paper:

Onyutha C, Willems P (2017) Space-time variability of extreme rainfall in the River Nile basin. International Journal of Climatology, DOI:10.1002/joc.5132

4.1 Introduction

High rainfall intensities are the drivers for floods, whereas dry conditions are of importance *e.g.* for agricultural applications. As already highlighted in Section 1.2, no clear results on spatio-temporal rainfall variability of especially extreme rainfall events in the Nile basin can be obtained from previous studies. This is because the previous studies based on short-term data of questionable quality tended to be confined to sub-basins or sub-regions while focusing on rainfall volumes instead of rainfall intensity extremes.

Therefore, the aim of this study was to investigate the variability in extreme rainfall using high-resolution ($0.5^\circ \times 0.5^\circ$) gridded long-term daily rainfall which spatially covers the entire region where the Nile basin is located. The spatio-temporal differences in the linkages of the rainfall variability across the study area to large-scale ocean-atmosphere interactions were also assessed.

4.2 Methodology

Daily series from 1948 to 2010 covering the entire Nile basin region (*i.e.* latitudes from 31° N to 5° S in the North–South direction, and longitudes over the range 24° E– 40° E in the West–East direction) were extracted from the global PGF-based data described in Section 2.2.1c.

For variability analyses, the following were extracted from the daily PGF-based data 1) the mean of ten Highest daily Rainfall Intensities in each year (HRI₁₀), 2) the Percentage of Dry Days in each year (PDD), and 3) the

longest period of dry days or the Maximum Dry Spell in each year (MDS). In this study, a dry day was considered as that with rainfall intensity below 1 mm/day.

To investigate the linkage of extreme rainfall variability to large-scale ocean-atmosphere conditions, several monthly climate indices (see Section 2.2.4) including AMO index, NAO index, IOD index, and Niño 3 index were considered. For variability analyses, the monthly series of the climate indices were converted to annual time scale to match that of the PDD, HRI₁₀ and MDS.

The QPM (already described in Section 3.2.1) was applied to derive anomalies for the HRI₁₀, PDD and MDS series at each grid point. Anomalies were also computed for the climate indices. Temporal variability of extreme rainfall series (HRI₁₀, PDD and MDS) along with climate indices were assessed using correlation analyses.

4.3 Results and discussion

4.3.1 Spatial distribution in dry conditions

Figure 4.1 shows, on average based on the data period 1948-2010, how dry conditions were distributed over the study area. The highest to the lowest PDD and MDS was obtained in the Northern, Central and Southern region respectively (Figure 4.1a-b). High values of the PDD and MDS show arid condition. The HRI₁₀ (though not provided in Figure 4.1) was found comparable with the map in Figure 2.2 except that the range of the HRI₁₀ was from 01 (in the Northern region) to 60 mm/day (for the Equatorial region).

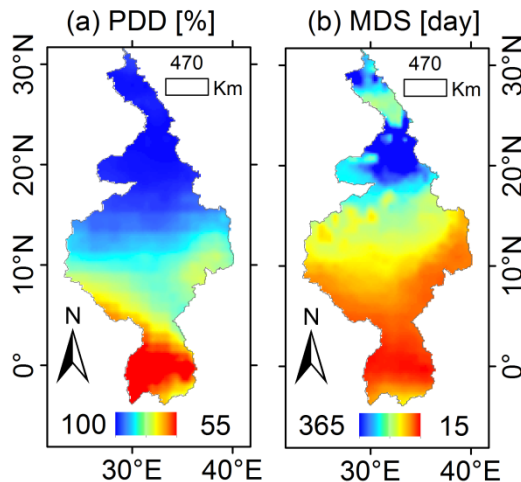


Figure 4.1 Spatial distribution of a) PDD, and b) MDS

4.3.2 Variability in the highest rainfall intensities

Figure 4.2 shows the spatial distribution of the correlation between variability in the EHRI and climate indices. It is noticeable that the HRI_{10} of the Equatorial region (or the Southern part) of the basin is linked to the IOD (Figure 4.2a), NAO (Figure 4.2b) and AMO (Figure 4.2c). These results are consistent with the findings from previous studies (Tierney *et al.*, 2013; Onyutha and Willems, 2015c) conducted for the areas including the Equatorial region. Some evidence is also visible on the linkage of the HRI_{10} to Niño 3 (Figure 4.2d). Generally, there are also several studies which document the linkage of rainfall variability in the Equatorial region to the ENSO (Schreck and Semazzi, 2004; Nicholson and Kim, 1997; Onyutha and Willems, 2015c; Phillips and McIntyre, 2000). The variation in HRI_{10} of the Central part of the study area is influenced by IOD, AMO and Niño 3 (Figure 4.2a, c-d). Mainly positive (negative) correlation was obtained between HRI_{10} and AMO (IOD and Niño 3). The variability of the annual maximum rainfall events observed over the upper Blue Nile basin located in the Central part of the study area was found mostly linked to the changes in SST or SLP in the Pacific and Atlantic Oceans (Taye and Willems, 2012). Jury (2010) reported that the rainfall variability in the Southern part of Ethiopia is influenced by the cold phase of the PDO as well as by the Southern meridional overturning circulation, whereas in the Northern

region it is enhanced by the warm phase of AMO. In Ethiopia, the Northern part is basically characterized by the Tropical Easterly jet and has a unimodal rainy season, whereas the rainfall in the Southern area exhibits a bimodal pattern (Jury, 2010). The rainfall variability for the Central part of the Nile basin was also evidenced to be linked to the ENSO (Abtew *et al.*, 2009; Seleshi and Zanke 2004; Onyutha and Willems, 2015c; Diro *et al.*, 2011; Seleshi and Demaree, 1995; Jury, 2010). For the Northern part of the basin, the possible drivers of the variability in rainfall extremes emanate from the Indian Ocean (Figure 4.2a) or Pacific Ocean (Figure 4.2d). Moreover, positive correlation also exists between HRI₁₀ and IOD, or HRI₁₀ and Niño 3.

Based on the spatial correlation map shown in Figure 4.2, the River Nile basin can be divided into three regions, namely Southern (5° S to 5° N), Central (5° N to 20° N), and Northern (20° N to 30° N). These regions, which are consistent with those found in a recent study by Onyutha and Willems (2015c), correspond to the spatially contrasting distribution of rainfall in response to the latitudinal migration of the ITCZ across the study area. The ITCZ tends to stay longer in the Southern than Central part of the study area and never reaches Egypt or the extreme North of Sudan, *i.e.* areas characterized by arid conditions. It is possible that the variation in extreme rainfall of the Northern region might be under the influence of the North African monsoon. However, the variation in the rainfall of the Southern region may be controlled by the changes in the Equatorial circulation. Furthermore, the spatial distribution of the rainfall in the Southern region of the study area is also influenced by regional topographical features (Onyutha *et al.*, 2016). In the Equatorial region, some of the regional features which may interactively modify the variation in rainfall include the water bodies (*e.g.* Lake Victoria, Lake Kyoga, Lake Albert, Lake Turkana) (Thiery *et al.*, 2015) and mountains (such as Mount Elgon, Mount Rwenzori, Mount Kenya, *etc.*). For instance, Thiery *et al.* (2015) found that by inducing circular airflow with over-lake convective inhibition during daytime and the reversed pattern at night, the Lake Victoria has a profound influence on atmospheric dynamics and stability with respect to the East African regional climate. In the Central region of

the study area, *e.g.* over Ethiopia, the distribution of rainfall is also controlled by the Ethiopian Highlands, which tend to receive complex pattern of rainfall following the blockage of the moist unstable Westerly flow of the Congo Air from reaching the coastal areas of Central region of the study area (Nicholson, 1996).

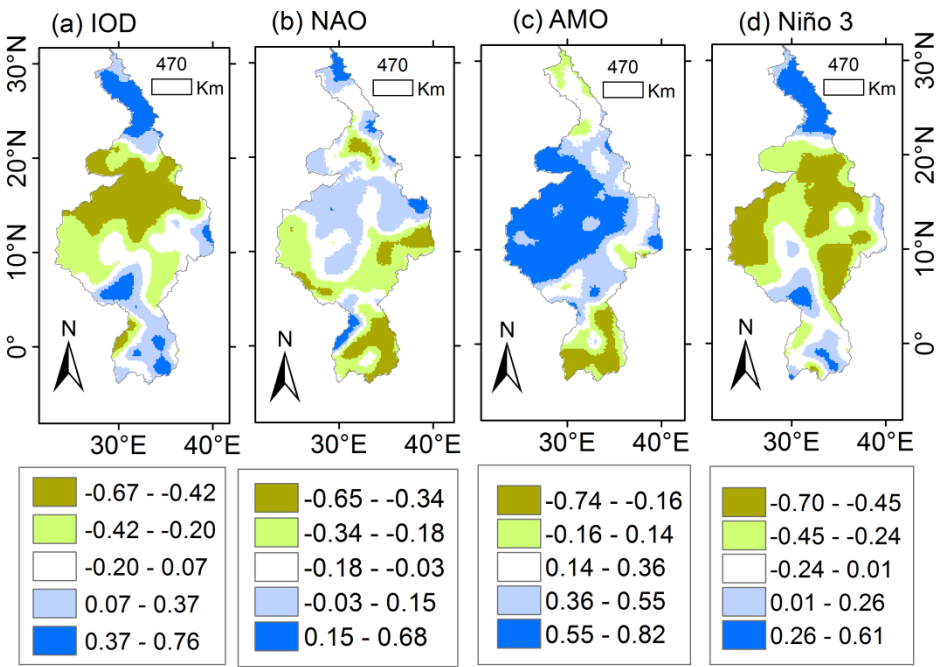


Figure 4.2 Correlation between QPM anomalies for the HRI₁₀ and a) IOD, b) NAO, c) AMO, and d) Niño 3

Figure 4.3 shows the temporal variability of HRI₁₀ along with climate indices. For the Equatorial region (Figure 4.3a), the HRI₁₀ oscillation highs (*i.e.* epochs of "above-normal" extreme rainfall events) were from the late 1950s to mid-1960s, and again in the late 1990s. On the other hand, oscillation lows (*i.e.* epochs of "below-normal" extreme rainfall events) were in the early 1950s and the 1980s. For the Central region (Figure 4.3b), the EHRI were above the reference in the 1950s and early 1960s. The extreme rainfall was below the reference from the late 1960s to the 1980s. Generally, the period from the late 1960s to the 1980s has been sufficiently documented for the decrease in rainfall in the Central part of the study area (Hulme, 1992; Zhang *et al.*, 2012; Onyutha and Willems, 2015c). For the

annual rainfall, this decrease in rainfall was about 30% (Hulme, 1992). Furthermore, this rainfall decrease led to much shorter (than normal) rainy seasons in the 1970s and the 1980s, mainly due to an earlier end. In the Northern part of the basin, the extreme rainfall events were below (above) their long-term average from the 1940s to the early 1980s (mid-1980s to the late 2000s) (Figure 4.3c). These results for all three regions are consistent with the findings from the spatio-temporal variability analyzed in a previous study (Onyutha and Willems, 2015c) based on observed annual rainfall totals.

To find out what percentage of variance in the rainfall extremes can be explained by the selected climate indices, Multiple Linear Regression (MLR) equation was used. The statistical "goodness-of-fit" metric coefficient of determination (R^2) was calculated for the various combinations of the climate indices as predictors. It is well-known that when more predictors are added to the MLR equation, the value of R^2 gets higher. Therefore, to attach some punitive measure to the addition of many variables, an adjusted R^2 (Ezekiel, 1930) was adopted as summarized in Table 4.1. The adjusted R^2 was optimized through minimization of the mean squared error. When all the four climate indices were used as predictors in the MLR equation, the adjusted R^2 obtained was 28.1%, 67.0%, and 70.3% for the Southern, Central and Northern regions, respectively. However, when the climate indices were used individually as predictors, the highest value of the adjusted R^2 was obtained using NAO (12.7%), AMO (50.1%) and IOD (29.2%) for the Southern, Central and Northern regions, respectively. For each region, the climate index with the highest adjusted R^2 was the best in describing the HRI_{10} variation. In other words, without considering the most suitable large-scale ocean-atmosphere condition, even if two or three climate indices were combined, the adjusted R^2 still remained low.

Generally, the percentage of the HRI_{10} that can be explained by the climate indices depends on the suitability of the oceanic areas from which the SLP or SST can resonate well with the variation in the extreme rainfall of the region under consideration. The influence of atmospheric circulation on

rainfall distribution over the various regions of the study area can be affected by the differences in the SLP from the different oceans. For the three regions of the study area, Onyutha and Willems (2015c) already investigated the co-variation of rainfall for the Southern, Central and Northern region of the study area with the SLP difference between the Indian and Atlantic oceans, as well as the Indian and Pacific oceans. Onyutha and Willems (2015c) showed that, for instance, the difference in the SLP taken over the Pacific and Indian Oceans is not significantly (at 5% level) correlated with the anomalies in the annual rainfall of the Southern region. The insignificance of this correlation is because, according to Tierney *et al.* (2013), it is mainly the Indian Ocean SST that drives East African rainfall variability by altering the local Walker circulation, whereas the influence of the Pacific Ocean is minimal. Instead, the variability in the annual rainfall of the Southern region was shown to be significantly (at 5% level) explained by the differences in SLP from the North Atlantic and Indian Oceans (Onyutha and Willems, 2015c). For the rainfall variability of the Central region of the study area, influences emanate from the Pacific and Indian Oceans. However, influences for the variability of rainfall in the Northern region come partially from the Indian, Atlantic and Pacific oceans.

Table 4.1 Adjusted R^2 (%) in predicting HRI₁₀ using climate indices

| Predictor | Region | | |
|------------|----------|---------|----------|
| | Southern | Central | Northern |
| IOD (I) | 0.40 | 12.7 | 29.2 |
| AMO (A) | 10.6 | 50.1 | 12.7 |
| NAO (N) | 12.7 | 2.90 | 0.30 |
| Niño 3 (E) | 1.50 | 20.8 | 14.1 |
| I, A | 10.8 | 59.6 | 44.8 |
| I, N | 20.7 | 29.6 | 44.0 |
| I, E | 1.50 | 24.9 | 32.7 |
| A, N | 19.7 | 50.2 | 14.2 |
| A, E | 11.0 | 60.9 | 33.0 |
| N, E | 12.9 | 21.2 | 14.2 |
| I, A, E | 11.0 | 65.0 | 51.5 |
| I, N, E | 23.3 | 31.9 | 44.1 |
| A, N, E | 19.7 | 61.0 | 33.0 |
| I, A, N | 25.2 | 65.4 | 70.3 |
| I, A, N, E | 28.1 | 67.0 | 70.3 |

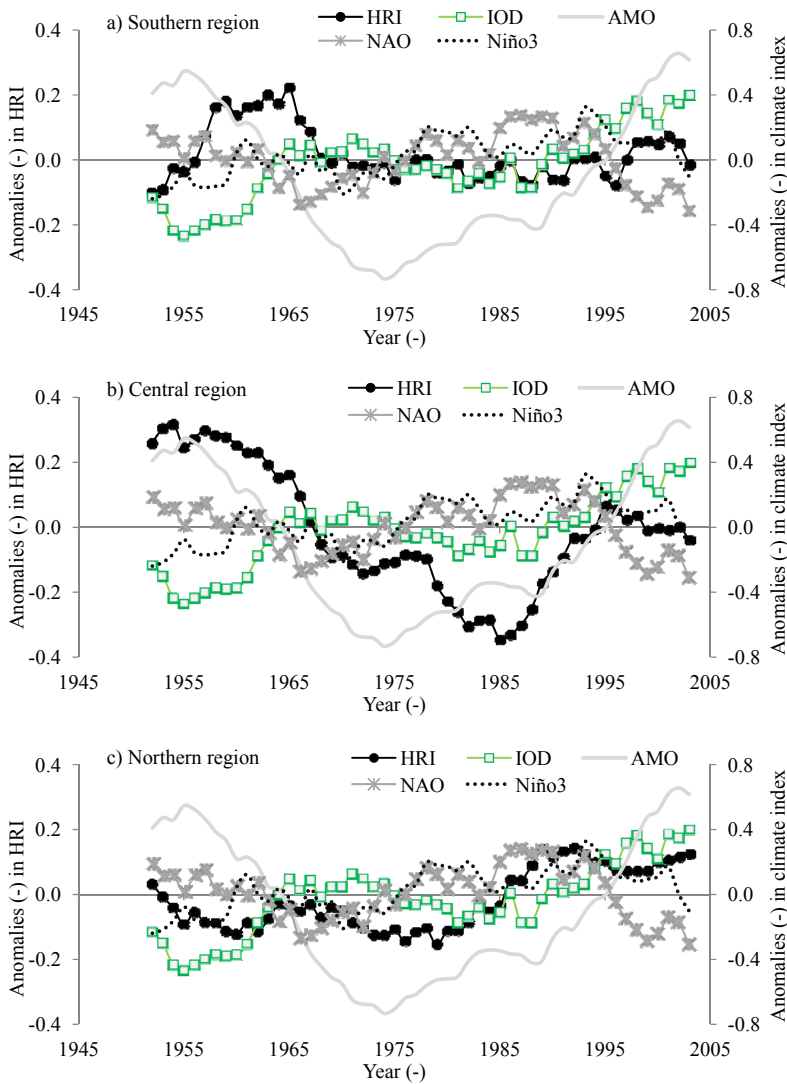


Figure 4.3 Variation in HRI₁₀ along with climate indices

4.3.3 *Variability in the lowest rainfall intensities*

a) Percentage of Dry Days (PDD)

Figure 4.4 shows how the PDD variability is linked to the large-scale ocean-atmosphere conditions. The PDD of the Northern extreme end of the study area is negatively correlated with IOD and Niño 3 (Figure 4.4a, d). The PDD

variability over the Southern region is positively correlated with AMO (Figure 4.4c). Over a large area of the Central part, the PDD variability is shown to be positively (negatively) correlated with the Niño 3 (AMO) (Figure 4.4c-d). Generally, the influence of climate indices on the PDD variability (Figure 4.4) is not as spatially coherent as that for the HRI₁₀ (Figure 4.4). This suggests that, apart from the influence from the large-scale ocean-atmosphere interactions, the dry conditions over the various regions of the study area may also be due to other drivers. Such drivers may include the influence from topographical features, spatial difference in the micro-climate, *etc.* Using a physically-based or detailed regional climate model, which can reasonably reproduce the hydro-climatic conditions across the study area, the influence of regional features such as water bodies, changes in land cover or use, *etc* on the rainfall distribution may be investigated.

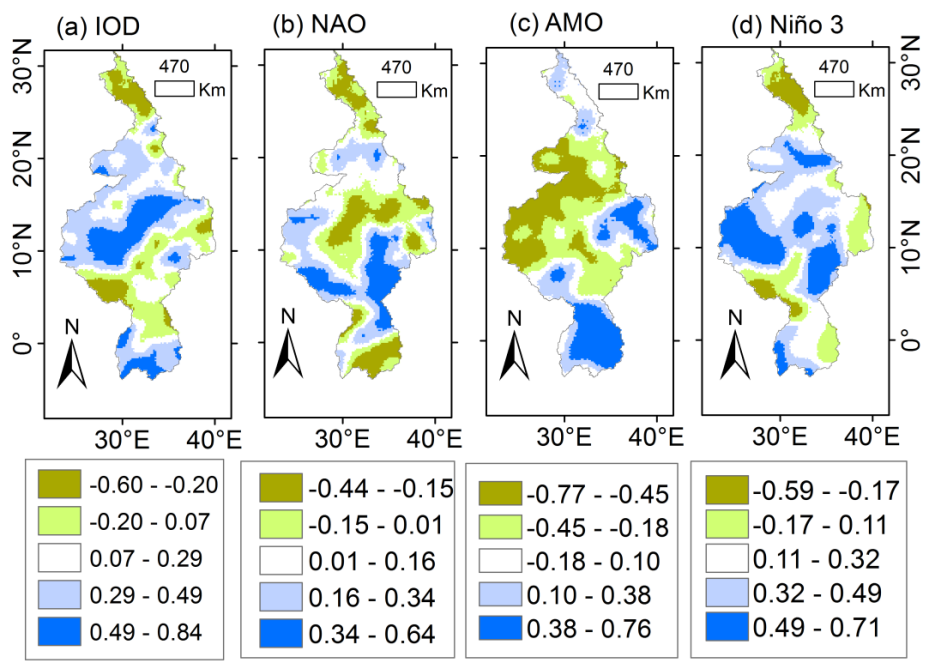


Figure 4.4 Correlation between QPM anomalies for the Percentage of Dry Days (PDD) in each year and a) IOD, b) NAO, c) AMO, and d) Niño 3

Figure 4.5 shows the temporal variability of the climate indices along with the averaged PDD over the considered regions. The PDD oscillation low (high) in the Southern region (Figure 4.5a) was from the mid-1950s to early 1970s (1980s to 2000s). For the Central region (Figure 4.5b), the PDD was above (below) the long-term mean from the late 1970s to early 2000s (1950s to mid-1960s). In the Northern part of the basin (Figure 4.5c), the PDD exhibited an oscillation low (high) from the early 1980s to the late 2000s (1950s and 1970s).

The statistical measures of the co-occurrence of the PDD and climate indices shown in (Figure 4.5) are summarized in Table 4.2. By combining all the four climate indices as predictors in the MLR equation, the adjusted R^2 was 41.5%, 68.6%, and 57.8% for the Southern, Central and Northern regions, respectively. When the climate indices were considered individually as the predictor, the highest value of the adjusted R^2 was obtained using the IOD (10.3%), Niño3 (35.3%) and IOD (21.5%) for the Southern, Central and Northern regions, respectively.

Table 4.2 Adjusted R^2 (%) in predicting PDD using climate indices

| Predictor | Region | | |
|------------|----------|---------|----------|
| | Southern | Central | Northern |
| IOD (I) | 10.3 | 20.7 | 21.5 |
| AMO (A) | 9.90 | 33.3 | 3.60 |
| NAO (N) | 1.60 | 3.80 | 3.00 |
| Niño 3 (E) | 7.40 | 35.3 | 11.7 |
| I, A | 21.7 | 50.7 | 26.4 |
| I, N | 22.0 | 45.7 | 44.4 |
| I, E | 13.0 | 41.6 | 24.9 |
| A, N | 13.5 | 34.1 | 8.00 |
| A, E | 21.3 | 57.7 | 18.4 |
| N, E | 7.90 | 35.6 | 12.6 |
| I, A, E | 26.9 | 64.1 | 31.5 |
| I, N, E | 22.0 | 50.8 | 45.2 |
| A, N, E | 22.9 | 55.1 | 20.2 |
| I, A, N | 41.5 | 62.1 | 57.3 |
| I, A, N, E | 41.5 | 68.6 | 57.8 |

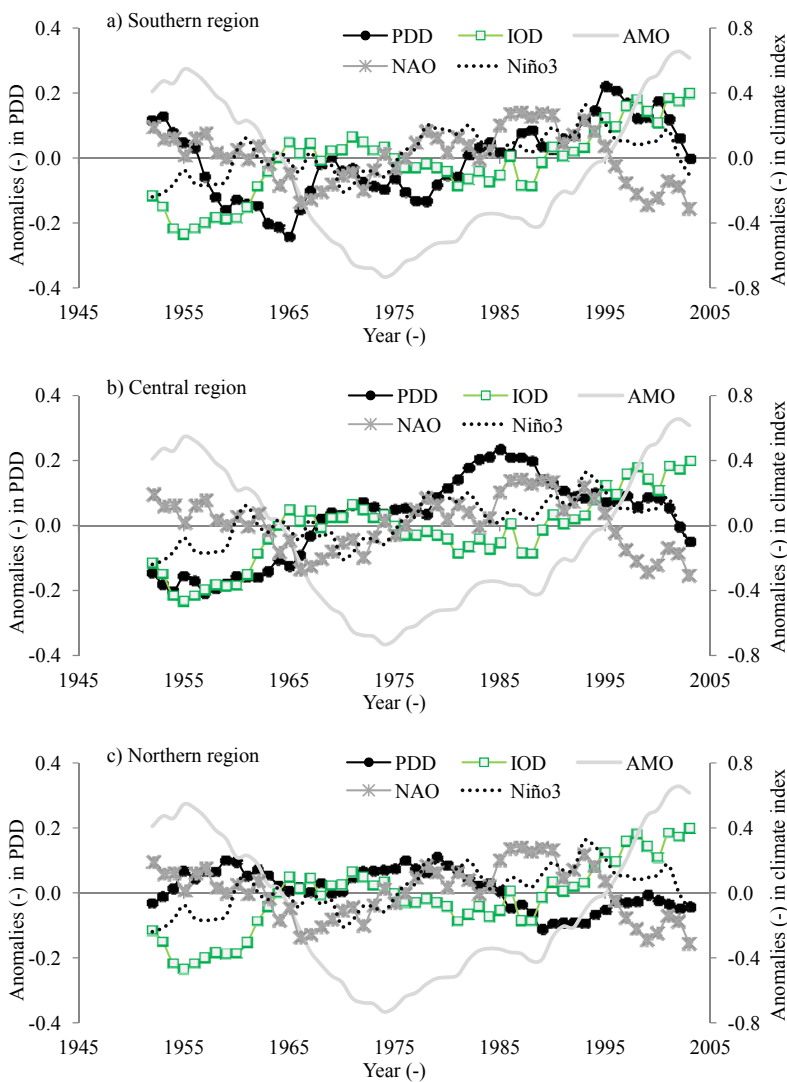


Figure 4.5 Temporal variation of PDD along with the climate indices

b) Maximum Dry Spell (MDS)

Figure 4.6 shows the strength of the linkage of the MDS anomalies to the climate indices. As seen from Figure 4.6d, the MDS of the extreme Northern part is negatively correlated with IOD and Niño 3. For the Eastern part of the Equatorial region, MDS is positively correlated with AMO (Figure 4.6c).

For the Central region, the correlation between MDS and climate indices (Figure 4.6a-d) is not as spatially coherent as that between HRI_{10} and the climate indices (Figure 4.2a-d). The variability in the dry conditions in the Central part of the basin, though partly explained by the anomalies in the climate indices, was found to be highly linked with the variability in the moisture flux (Long *et al.*, 2000; Zhang *et al.*, 2012). Recently, Zhang *et al.* (2012) found that drought events in this part of the Nile basin result from the tendency of the Northeasterly wind in Sudan, which limits Northward propagation of moisture flux to the Northern part of Sudan (and eventually Egypt). Based on the analysis of atmospheric water and moisture fluxes in West African Monsoon, the rainfall during the dry season in the Sahel region (which is located in the Central part of the basin) was found to be clearly dominated by the Easterly anomalies in the moisture flux South of 15° N (Fontaine *et al.*, 2003).

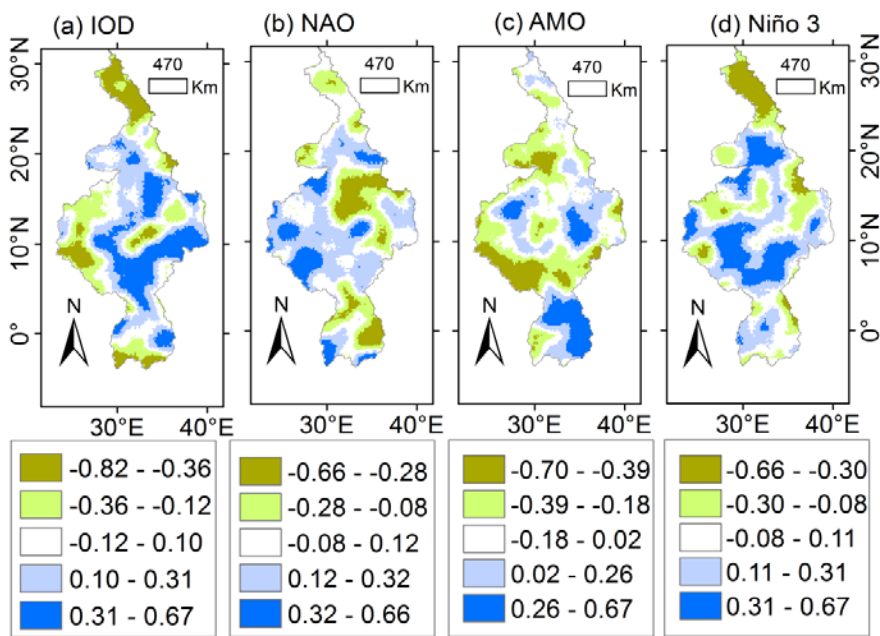


Figure 4.6 Correlation between QPM anomalies for the Maximum Dry Spell (MDS) in each year and a) IOD, b) NAO, c) AMO, and d) Niño 3

The temporal variation in the MDS along with climate indices is shown in Figure 4.7. The MDS of the Northern part of the basin (Figure 4.7c) was

above (below) the reference in the 1950s to 1970s (mid-1980s to around 2000). For the Central region (Figure 4.7b), the MDS exhibited oscillation lows (high) in the 1950s and 1990s (from the late 1960s to around 1990). For the Equatorial region (Figure 4.7a), the MDS was below (above) the reference in the mid-1990s (from the mid-1960s to late 1970s).

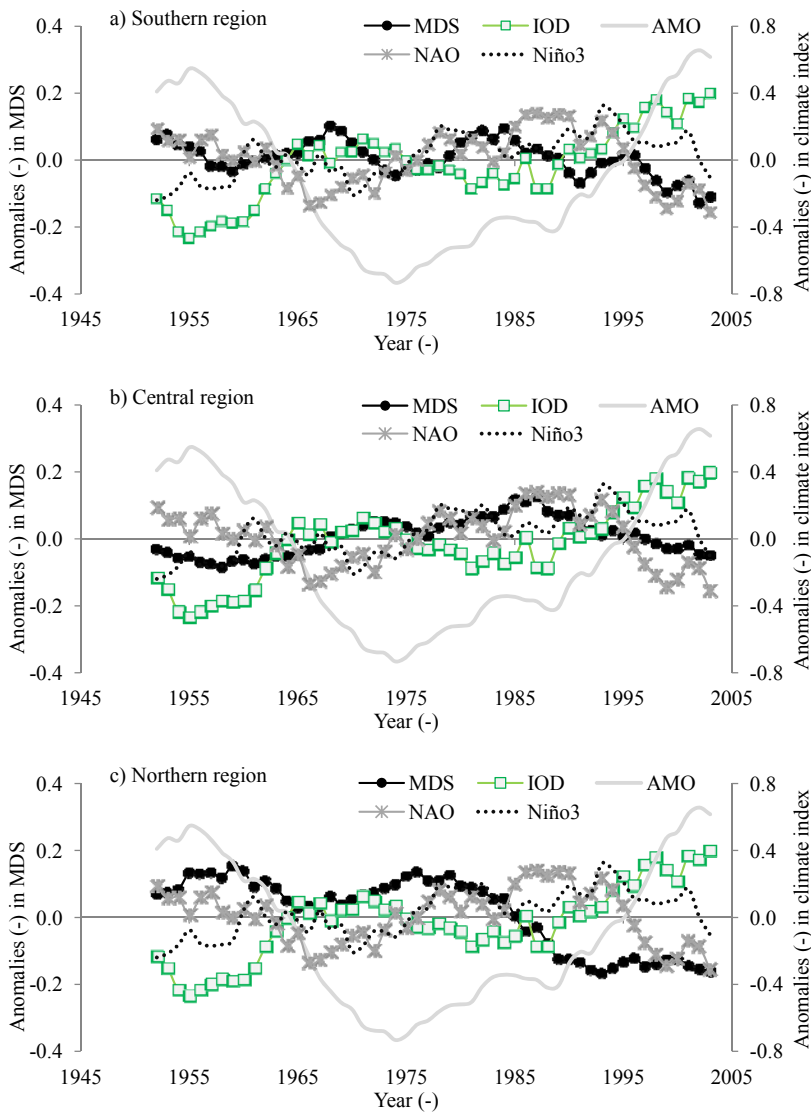


Figure 4.7 Temporal variation in the maximum dry spell (MDS) in each year along with climate indices

The strength of the linkage of MDS variability to climate indices shown in Figure 4.7 is summarized in Table 4.3. By combining all the four climate indices as predictors in the MLR equation, the adjusted R^2 was 38.5%, 73.9%, and 80.5% for the Southern, Central and Northern regions, respectively. When the climate indices were considered individually, the largest percentage of MDS variance was found to be explained by IOD (21.8%), AMO (54.7%) and IOD (52.0%) for the Southern, Central and Northern regions, respectively.

It is possible that the linkage of the variation in extreme rainfall metrics such as HRI_{10} , PDD and MDS to the large-scale ocean-atmosphere conditions may be influenced by other factors. For instance, Onyutha and Willems (2017) recently showed that due to the increase in the heterogeneity of rainfall over the study area, as the size of the spatial domain for analysis increases, the climate indices explain rainfall variability more suitably at a regional than location-specific spatial scale. Furthermore, for a region that is not perfectly homogenous as the study area, the occurrences of extreme rainfall conditions can be influenced in a synergistic way by the variation of two or more possible drivers. For instance, the variation of extreme rainfall of a particular region can be linked to both of the ENSO episodes including the warm (El Niño) and cold (La Niña) phases, which are known to bring about conditions that are respectively wetter and drier than the normal one.

Table 4.3 Adjusted R^2 (%) in predicting MDS using climate indices

| Predictor | Region | | |
|------------|----------|---------|----------|
| | Southern | Central | Northern |
| IOD (I) | 21.8 | 3.60 | 52.0 |
| AMO (A) | 12.8 | 54.7 | 6.00 |
| NAO (N) | 5.30 | 16.8 | 0.60 |
| Niño 3 (E) | 2.60 | 15.1 | 25.1 |
| I, A | 37.1 | 56.7 | 60.7 |
| I, N | 21.9 | 38.6 | 63.7 |
| I, E | 21.9 | 15.3 | 58.2 |
| A, N | 15.6 | 62.2 | 6.10 |
| A, E | 18.2 | 61.1 | 37.0 |
| N, E | 10.2 | 25.8 | 29.1 |
| I, A, E | 37.4 | 61.4 | 70.1 |
| I, N, E | 21.9 | 38.7 | 64.4 |
| A, N, E | 23.0 | 66.3 | 39.2 |
| I, A, N | 38.5 | 73.9 | 79.6 |
| I, A, N, E | 38.5 | 73.9 | 80.5 |

4.3.4 Significance of the temporal variability

Figure 4.8 shows the spatial distribution of grid points at which the QPM anomalies of HRI₁₀, PDD and MDS exhibited oscillation high (OH) or low (OL) periods significant at the 5% level. For the variation in the HRI₁₀ and PDD, there were more grid points with significant OH and OL periods in the Central than the Southern and Northern regions (Figure 4.8a-b, d-e). The MDS exhibited more significant OLs than OHs (Figure 4.8c, f). Significant OHs of PDD and MDS indicate severe drought events. Due to prolonged and severe dry conditions, some areas within the study area experience famine or food insecurity due to failure of rain-fed crops, and death of animals. Dry conditions also affect the flow control for ecological purposes such as compensation flow and dilution flow for improving the water quality for power generating plants. Significant OHs of HRI₁₀ lead to floods, landslides, *etc*, which are not exceptional to the study area as highlighted in Section 1.1.

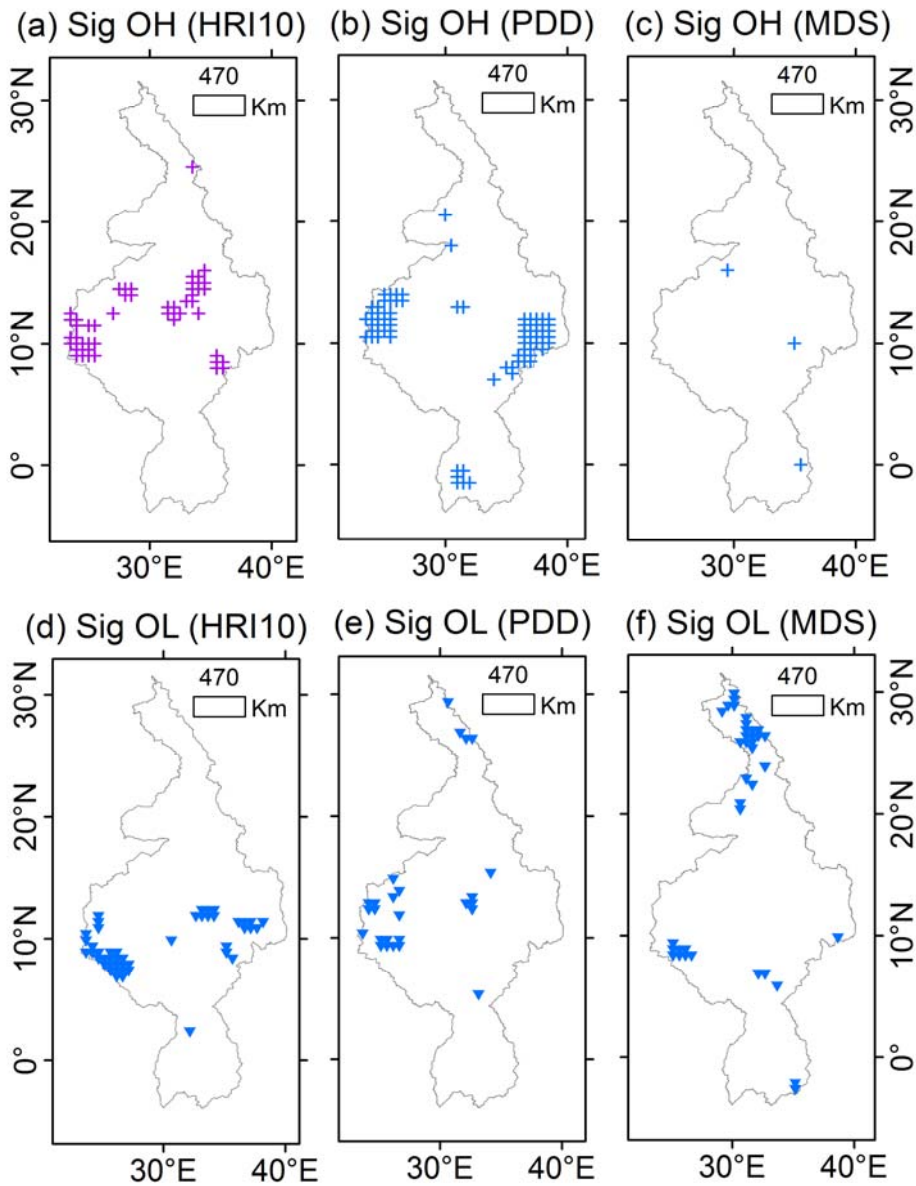


Figure 4.8 Locations with QPM anomaly significant (sig) at 5% level for the (a, d) HRI₁₀, (b, e) PDD, and (c, f) MDS over the period 1948-2010; the OH and OL denote positive and negative anomalies respectively

4.4 Conclusions and relevance of the findings

4.4.1 Answers to research questions

Is there variability in extreme rainfall conditions? Is the variability of dry conditions and EHRI significant? Can the variability of the extreme rainfall be explained in terms of the variation in large-scale ocean-atmosphere conditions?

Based on a 15-year moving window, it was found that the EHRI and dry conditions show oscillatory behavior over multi-decadal time scales. Oscillation highs and lows in the dry and wet conditions were more significant at 5% level in the Central than the Southern and Northern parts of the study area. The amount of variability in HRI₁₀ or dry conditions that could be explained by climate indices when considered individually was up to more than 50%.

4.4.2 Relevance of the findings

These findings are vital for water management and planning. Given the data scarcity in the study area, quantiles estimated using short-term series can be biased. The use of biased quantiles for risk-based applications can lead to serious cost implications on the one hand, and safety problem on the other. To minimize the bias, the linkage between the EHRI and climate indices can be used to obtain improved quantiles.

When the driver of the variability is known, an upcoming OH or OL period can be predicted. The EHRI may occur more frequently during periods of OHs than those of the OLs. By taking into account the variation in the EHRI frequency over the OHs and OLs, considerate evacuation measures or resettlement plan can be put in place for people who reside in flood- or landslide-prone areas. If economically feasible, large water detention basin could be excavated along river tributaries to store the surplus water (for some time) to protect against flooding. Hereby, the potential clustering of water surplus or shortage periods should be accounted for.

For periods of "below-normal" rainfall conditions, more dry than wet conditions may occur. Alternative sources of power supply could be planned for the population, again taking the new insights in the multi-decadal variations into account. Low water quality during dry conditions

deteriorates power plants (*e.g.* through scaling or corrosion of turbines) thereby shortening their design life. Expensive repair of power plants can be avoided if possible *e.g.* through careful planning considering the frequency of EHRI during periods of OLs.

Chapter Five

Influence of spatio-temporal scales on variability analyses

This chapter is based on the paper:

Onyutha C, Willems P (2017) Influence of spatial and temporal scales on statistical analyses of rainfall variability in the River Nile basin. Dynamics of Atmospheres and Oceans, 77:26–42

5.1 Introduction

For the different regions of the Nile basin, several studies have been conducted on the teleconnection of rainfall variability to large-scale ocean-atmosphere interactions (Nicholson and Entekhabi, 1986; Nyeko-Ogiramoi *et al.*, 2013; Onyutha and Willems, 2015c; Phillips and McIntyre, 2000; Taye and Willems, 2012; Tierney *et al.*, 2013). In a study relevant for agricultural practices, Phillips and McIntyre (2000) investigated the ENSO and inter-annual rainfall variability in Uganda. Using station-based observed rainfall at various locations across the entire study area, analyses of the variability in seasonal and annual rainfall were also conducted by Onyutha and Willems (2015c) based on the quantile perturbation approach. The authors used series with record lengths of 40 years and above to investigate the causes of rainfall variability with respect to the global SST and SLP as well as 10 climate indices. Taye and Willems (2012) and Nyeko-Ogiramoi *et al.* (2013) applied QPM to analyze the variation in rainfall extremes of the Blue Nile basin and LVB respectively. These authors also assessed the co-occurrence of the anomalies in the rainfall extremes and those of some climate indices such as the IOD, AMO, PDO and NAO index. Tierney *et al.* (2013) applied the EOF to investigate the influence of the Indian Ocean on the multi-decadal variability in East African hydro-climate. From the above studies, the selection of the following factors for analyses tended to differ: the variability computation method, spatial coverage, temporal scale, rainfall variability driver, *etc.* Generally, the above factors as well as the questionable data quality (*e.g.* missing records), data limitation (*e.g.* short-term records), *etc.* influence the analyses of rainfall variability and its driving influence. Nonetheless, in all the above cited studies for the study

area, little or no attention was given to investigate how rainfall variability analyses are influenced by any of the several aforementioned factors. It may be possible that the different factors may influence rainfall variability synergistically. However, for an in-depth investigation of their influence on rainfall variability, one factor can be assessed at a time. For instance, in this study only the variation in the spatio-temporal scales was considered.

Some of the matrix methods for statistical analyses of structures in large datasets (as considered in this study) include the EOF, Maximum Covariance Analysis (MCA), Canonical Correlation Analysis (CCA), *etc.* As opposed to the use of space and time dimensions by the EOF to explain the maximum amount of variability in a given data matrix, the MCA and CCA consider two matrices with the same sampling dimension but of different structures. The application of the MCA especially for meteorological and oceanographic studies follows its popularization by Wallace *et al.* (1992) and Bretherton *et al.* (1992). However, it is vital to note the extreme caution by Cherry (1996) that both the MCA and CCA tend to produce spurious spatial patterns which may not be physically meaningful or explainable. Eventually, the EOF was adopted for the analyses in this study. Moreover, the robustness of the EOF method for analyses of rainfall variability in the study area was already demonstrated in some studies such as Indeje *et al.* (2000), Ogallo (1988), Tierney *et al.* (2013), and Yeshanew and Jury (2007).

Therefore, this study was specifically aimed at investigating the influence of spatial and temporal scales on the analyses rainfall variability. This was done while putting rainfall variability drivers into perspective.

5.2 Methodology

5.2.1 Extraction of data for rainfall variability analysis

To extract PGF-based rainfall series described in Section 2.2.1, spatial domains were selected in an increasing size from $1^{\circ} \times 1^{\circ}$ to $10^{\circ} \times 10^{\circ}$ (see Figure 5.1) such that the location of the point rainfall fell at the center of the area covered by each spatial domain. The rainfall series of the various sizes of spatial domains were extracted over the various regions of the

River Nile riparian countries (Figure 5.2). The location of point rainfall was marked in each of the selected countries or regions labeled (1) to (6) in Figure 5.2. Because the PGF data were of $0.5^{\circ} \times 0.5^{\circ}$ grid resolution, the total number of grid points from $h^{\circ} \times h^{\circ}$ size of spatial domain was $(2h+1)^2$. For instance, there were $(2 \times 10 + 1)^2 = 441$ grid points for the $10^{\circ} \times 10^{\circ}$ spatial domain of each region. Point rainfall at the center of the $1^{\circ} \times 1^{\circ}$ spatial domain was also considered. The daily rainfall at each grid point was converted to monthly, seasonal and annual time scales. For the seasonal time scale, only the MAM series (for regions 1-2), and JJAS rainfall (for regions 3-5) were considered.

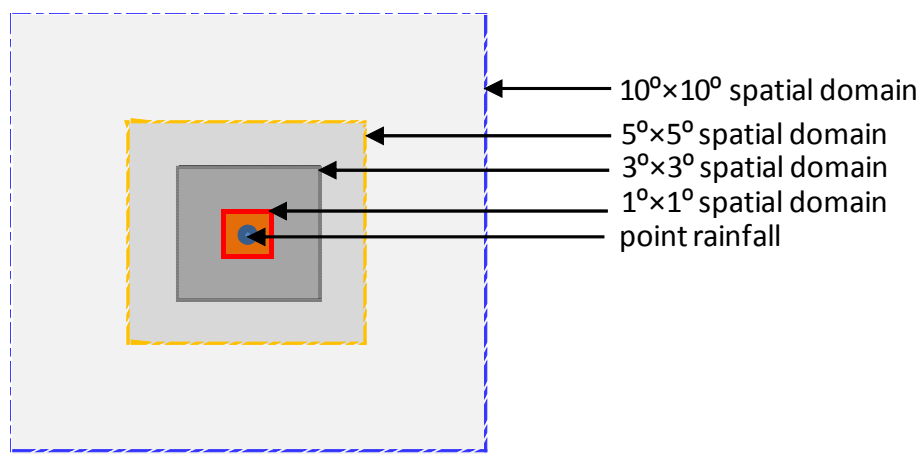


Figure 5.1 Various sizes of the spatial domains of the rainfall series considered for analyses

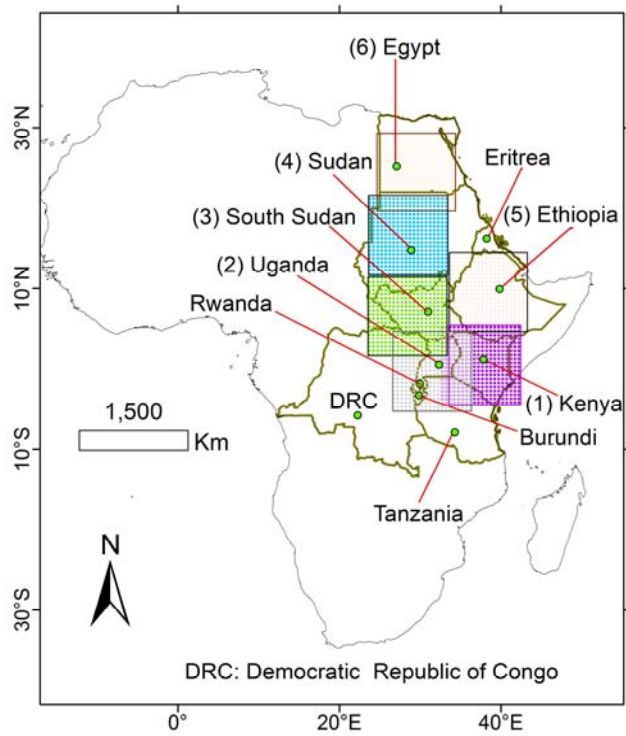


Figure 5.2 The River Nile riparian countries. The number in “()” denotes this study’s ID of the country in which the center of the spatial domain of size $10^{\circ}\times 10^{\circ}$ for the gridded rainfall is located

5.2.2 EOF analyses of rainfall variability

The EOF Analysis was performed considering the gridded rainfall:

- a) over the entire Nile basin as the spatial domain of the series,
- b) of various sizes of spatial domain (Figure 5.1) for each region, and
- c) of different temporal resolutions (*i.e.* monthly, seasonal and annual time scales) of each selected size of the spatial domain in each region shown (Figure 5.2).

Daily data may comprise a lot of white noise. On the other hand, in monthly data, the effect of seasonality (if strong) may be dominant. Therefore, for clarity of variability results, daily time scale was not considered, and

seasonal time variation was first removed from the monthly series before conducting the EOF analyses.

A small number of linearly uncorrelated series can be derived from the given inter-correlated datasets so as to explain the maximum amount of variance. The structures explaining the maximum amount of variance in the two dimensional series (*i.e.* gridded rainfall in this case) can be analyzed using the EOF. In the EOF analyses, space and time are used as the structure and sampling dimensions, respectively. The set of structures produced in the first (*i.e.* the structure) dimension are called the EOFs. In the sampling (*i.e.* time) dimension, Principal Components (PCs) are produced as the complementary set of structures. The PCs and EOF's are orthogonal in their own dimension.

Consider that m variables Y_1, Y_2, \dots, Y_m are correlated with each other and Q is to be predicted given the Y_j 's. Before predicting Q , it would be valuable to establish a new predictor Z as linear combinations of the Y_j 's using:

$$\left. \begin{aligned} Z_1 &= a_{11}Y_1 + a_{12}Y_2 + \dots + a_{1j}Y_j \dots + a_{1m}Y_m \\ Z_2 &= a_{21}Y_1 + a_{22}Y_2 + \dots + a_{2j}Y_j \dots + a_{2m}Y_m \\ &\dots\dots\dots \\ Z_i &= a_{i1}Y_1 + a_{i2}Y_2 + \dots + a_{ij}Y_j \dots + a_{im}Y_m \\ &\dots\dots\dots \\ Z_n &= a_{n1}Y_1 + a_{n2}Y_2 + \dots + a_{nj}Y_j \dots + a_{nm}Y_m \end{aligned} \right\} \quad (5.1)$$

In this study, Y_j is the rainfall series at the j^{th} grid point, N is the possible maximum number of dimensions to explain the desired amount of variance in the data, m equals to the number of grid points. In summary, $Z_i = a_{ij}Y_j$ such that the matrix of coefficients a_{ij} (for the i^{th} observation at the j^{th} grid point) rotates the original set of series into a second one. The values of the a_{ij} can be determined such that the Z_i 's are uncorrelated and orthogonal variables. Eventually, Z_1 explains the maximum amount of variance of the Y_j 's, the Z_2 explains the maximum amount of the remaining variance of the Y_j 's, *etc.* The values of the coefficients a_{ij} are the EOFs onto which when the

series Y_j are projected, the PCs *i.e.* Z_i (the original series in the new coordinate system) are obtained. In other words, the coefficients a_{ij} are the “loadings” and represent the weights of the original variables in the PCs. It is the orthogonality (*i.e.* lack of correlation in time) which makes the PCs very efficient and suitable for analyses of variability.

The extraction of the PCs in this study was based on the correlation matrix such that the n order eigenvector (a_{11} a_{12} , ..., a_{1N}) included the coefficients of N PCs. Other bases of PC extraction include variance, covariance, *etc.* The amount of variability explained by the first, second, ..., N^{th} EOF is referred to as the first (η_1), second (η_2), ..., N^{th} (η_N) eigenvalues, respectively such that $\eta_1 \geq \eta_2 \geq \dots \geq \eta_N$. The proportion of the variance explained by the i^{th} component is expressed as $\eta_i / \sum \eta_i$.

Stable patterns in the form of localized structures are expected to emerge over the spatial domain of analysis. However, this expectation is affected by the significant amplitude which tends to occur due to effect of the orthogonality constraint. Eventually, to reduce the effect of orthogonality constraint and achieve a reasonably stable spatial patterns, rotation of the eigenvectors can be considered (Horel, 1984; Richman, 1986). Based on the rotation, regions with similar temporal variations are isolated thereby improving the identification of regions with maximum correlation between the variables and the components (Santos *et al.*, 2010). Several methods exist for rotation of the eigenvectors including the Varimax, Quartimax, Equamax, Parsimax, Quartimin, Promax, Orthomax, Oblimin, *etc.* In this study, the Varimax procedure which is the most commonly used was adopted because according to Richman (1986) it preserves the orthogonality, and gives more physically explainable variability patterns than other approaches. Varimax is also known to exhibit less sensitivity to the number of variables in the EOF analysis. Varimax method employs the approximation of simple structure of individual factors *i.e.* the variance of its squared loadings. The sum of the variance values is also referred to as the ‘row Varimax criterion’ in which large eigenvalues (*i.e.* those with large factor loadings) contribute more to the criterion than those with small

factor loadings. Eventually, the bias on the first eigenvector can be reduced using the weights of the variables based on their communalities.

5.2.3 Correlation analysis

Under the H_0 of no correlation between rainfall variability and climate indices, correlation analyses was conducted at the significance level α_s of 5%. Figure 5.3 shows, for illustration, the co-occurrence of the climate indices and the factor scores for $10^\circ \times 10^\circ$ gridded annual rainfall in Kenya and Ethiopia. Since the magnitudes of the correlation coefficients of 0.55 (for Kenya, Figure 5.3a) and -0.63 (for Ethiopia, Figure 5.3b) were greater than the critical value of 0.25, the H_0 was rejected at the significance level α_s of 5% in both regions. The correlation analyses were performed in two steps. Firstly, considering the gridded rainfall over the entire Nile basin, and secondly, based on various sizes of the rainfall grids. In each of the two steps, correlation analyses were conducted for monthly, seasonal and annual time scales.

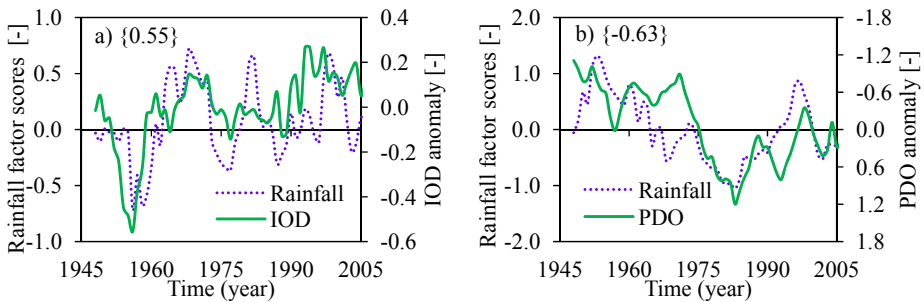


Figure 5.3 Variability in climate indices (IOD or PDO) and the factor scores obtained from the annual rainfall based on $10^\circ \times 10^\circ$ spatial domain with its center located in a) Kenya, and b) Ethiopia; the label in “{ }” shows the coefficient of correlation between rainfall and IOD or PDO

5.3 Results and discussion

5.3.1 Explained variance versus spatial scale

Figure 5.4 shows, for the annual rainfall, variation of the amount of explained variability with PCs. For all the time scales (though only shown

for annual rainfall), the explained variability was found to be small for higher than lower PCs. Considering the $1^\circ \times 1^\circ$ spatial domain, the variability explained by only the first PC ranged from about 50% (Figure 5.4a, c, e) to nearly 60% (Figure 5.4b, d, f). However, by combining the first and the second PCs, the cumulative variability explained in the annual rainfall went above 90% for the $1^\circ \times 1^\circ$ gridded rainfall of all the selected regions. Generally, as the size of the spatial domain was increased, the amount of explained rainfall variability noticeably reduced. This, which was clearly evident for the most important (*i.e.* first two) PCs, could be explained in terms of the increase in rainfall heterogeneity as the size of the spatial domain increased. As the distance between any two grid points increased, the magnitude of the correlation between the data at the selected locations (or grid points) reduced. Eventually, the average of the correlation was higher between data at the various grid points within a small- than that for large-sized spatial domain. This follows the influence of the micro-climate or features *e.g.*, water bodies, topography, *etc* which become more evident on the rainfall variation as the size spatial domain increases. Of course, it is known that the amount of variability that can be explained in the EOF depends on how highly correlated the variables for the analyses are. As the size of the spatial domain increases, the heterogeneity in the rainfall also increases. In the same line, as the rainfall series at the various grid points within a selected size of spatial domain become less correlated, the amount of variability explained in the EOF also reduces. Eventually, from $1^\circ \times 1^\circ$ to $10^\circ \times 10^\circ$ size of the spatial domain, the cumulative variability which could be explained from the first two PCs dropped by 43.8, 53.2, 55.1, 36.7, 56.8, 69.1% for regions 1 to 6 as shown in Figure 5.4a-f respectively.

5.3.2 Explained variance versus temporal scale

Figure 5.5 shows the influence of the variation of temporal resolution on rainfall variability considering a given size of spatial domain. Generally, as the rainfall size of the spatial domain increased, the amount of explained variability reduced as already highlighted in Figure 5.4. For a selected size of spatial domain, the amount of explained variability reduced as the temporal resolution of the rainfall increased *i.e.* from annual to seasonal, and monthly time scale (Figure 5.5a-f). This was because of the temporal

smoothing of series through aggregation *e.g.* from monthly to annual which removed the short-term fluctuations to obtain a more representative data and allowed the study of its general behavior. This realization was expected based on Figure 2.5 in which the measure of variability was demonstrated to vary from the highest to the lowest in monthly, seasonal and annual time scales, respectively.

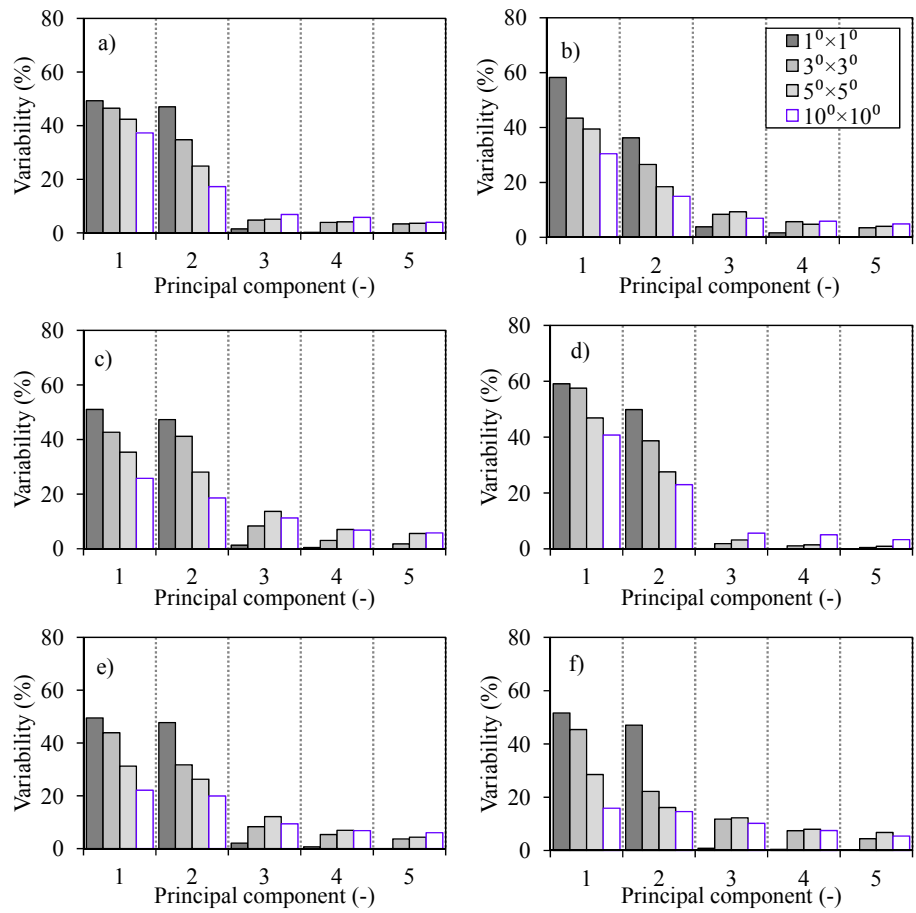


Figure 5.4 Percentage of variability explained in the annual rainfall from region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6; the charts share the same legend as in b) and the numbering of the regions is consistent with the IDs assigned on Figure 5.2

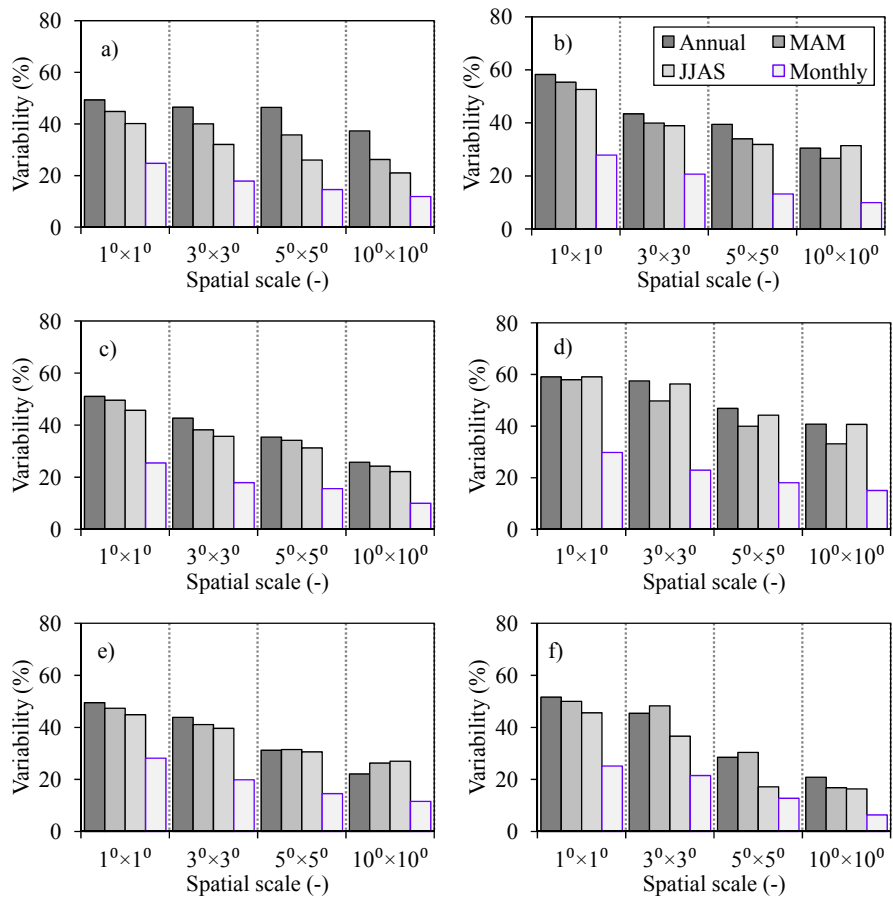


Figure 5.5 Percentage of variability explained by the first PC based on the rainfall for region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6; the charts share the same legend as in b)

5.3.3 Spatial and temporal variability

Figure 5.6 shows, for the entire Nile basin, the spatial variation of the leading EOF factor loadings. Whereas the grid points in the Equatorial region and those in the Northern part of the study area loaded somewhat negatively, the Central portion had positive EOF loadings for all the time scales (Figure 5.6a-d). The EOF loadings appear to divide the rainfall of the Nile basin into three parts *i.e.* that in the Southern, Central and Northern areas. This is highly consistent with the grouping of meteorological or rainfall stations in the study area from the recent study by Onyutha and

Willems (2015c). The variation in the EOF loadings across the study area suggests the regional differences in the rainfall variability driving forces. Such regional differences in the rainfall variation can be influenced by the micro-climate or features such as topography, water bodies, changes in land-use cover, *etc.* Moreover, the drivers of the rainfall variation may also differ from one time scale to another. Further investigation of the regional difference was made in a spatio-temporal way by considering the various regions of the study area. The overall distance between the equator and the areas with high positive factor loadings was found longer for the JJAS (Figure 5.6b) than that of the MAM (Figure 5.6c) season. This was because of the latitudinal migration of the ITCZ as already mentioned in Section 2.1.

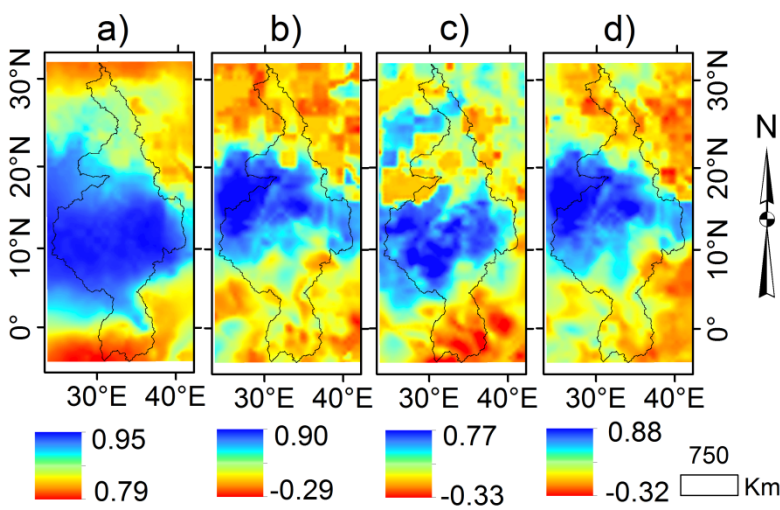


Figure 5.6 Leading EOF factor loadings after Varimax rotation for $0.5^{\circ}\times0.5^{\circ}$ gridded series extracted over the spatial domain covering the entire Nile basin based on a) monthly, b) JJAS, c) MAM, and d) annual rainfall

Figure 5.7 and Figure 5.8 respectively show the spatial view of the leading EOF factor loadings and the first PC. Locations with positive/negative loadings (Figure 5.7) show periods when the rainfall was above/below the reference or the mean of the long-term series as presented in Figure 5.8. For regions 1 and 2, the rainfall was below the reference in the 1950s, 1970s and the early 1990s; however, wet conditions occurred in the 1960s, 1980s and the late 1990s (Figure 5.8 a, d). These results agree with those of

Onyutha and Willems (2015c) who found that the rainfall in the Equatorial region was below (above) reference from the late 1940s to 1950s (1960s to mid-1980s). For regions 3 and 5, dry conditions occurred from the late 1970s to the early 1990s, and rainfall was above the reference from the early 1950s to late 1960s, and around 2000 (Figure 5.8 b). Using the maximum rainfall in each year, Taye and Willems (2012) obtained similar patterns of variability for the upper Blue Nile basin in Ethiopia. For regions 4 and 6 (arid condition), rainfall totals dropped below the reference from the late 1960s to the mid-1970s, and again from the early 1990s to mid-2000s. Somewhat wet periods occurred from the early 1950s to early 1960s, and again around 1980 (Figure 5.8c). For the Sahel (*i.e.* region 3), the decline in rainfall from the late 1960s to the mid-1970s was also found by Hulme (1992).

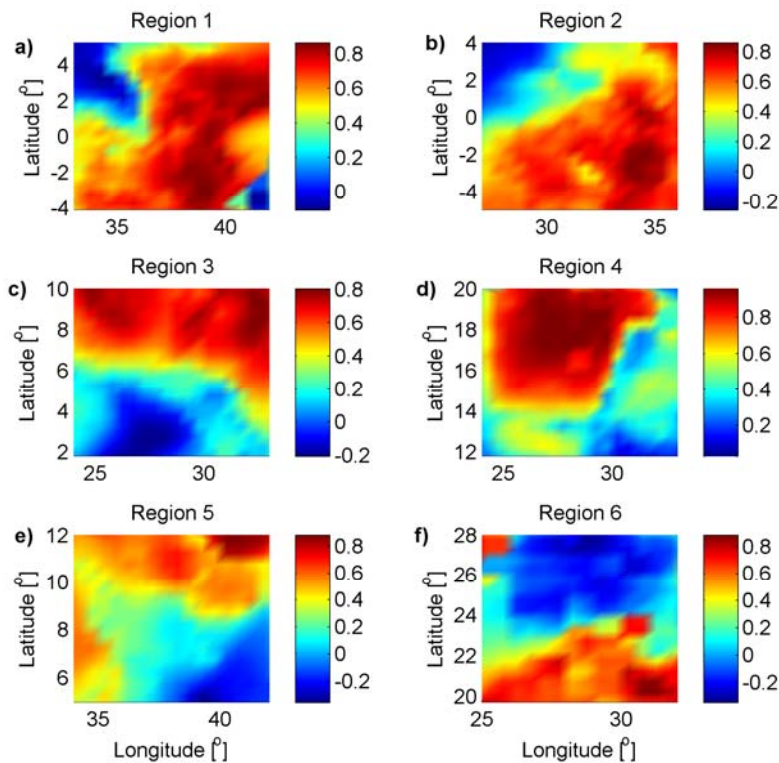


Figure 5.7 Leading EOF factor loadings for annual rainfall from the various grid points of the 10°×10° size of spatial domain from region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6. Label of horizontal axis for a)-d) is similar to that of e).

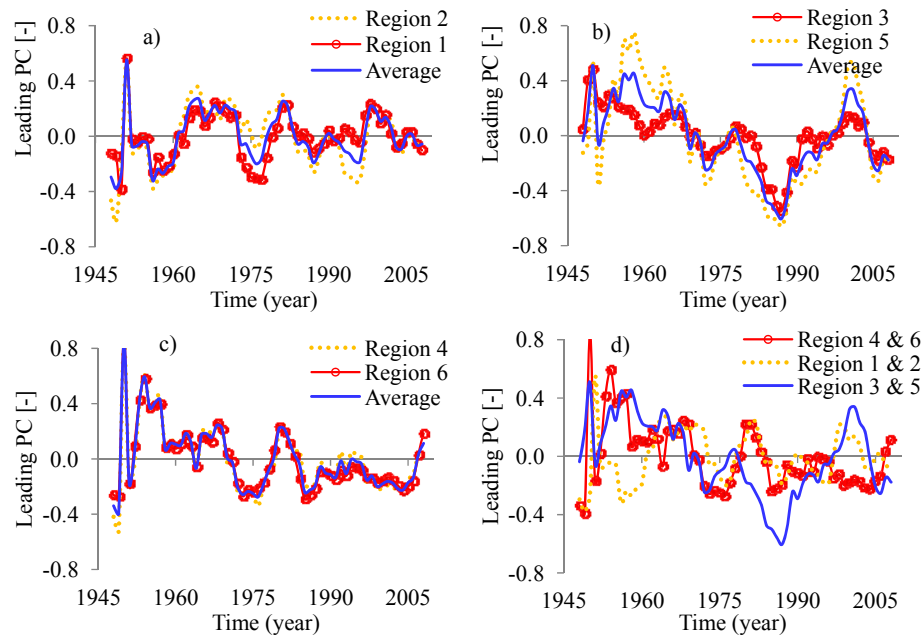


Figure 5.8 Leading PC based on $10^{\circ}\times10^{\circ}$ gridded annual rainfall of regions a) 1 and 2, b) 3 and 4, and c) 4 and 6; chart d) is for the average loadings of the various regions

5.3.4 Correlation analysis

Figure 5.9 and Figure 5.10 show the correlation between rainfall and climate indices. The regional difference in the rainfall variability across the study area as highlighted from Figure 5.6 is also demonstrated in Figure 5.9 and Figure 5.10. It is noticeable that some difference exist among the spatial correlation maps of the various time scales (Figure 5.9a-h and Figure 5.10a-h). Because of the tendency of the correlation to differ from one time scale to another, the climate indices IOD (for region 1 and 2), PDO (for region 3 and 5), and AMO (for region 4 and 6) were selected based on the spatial map of annual rainfall shown in Figure 5.9d, h and Figure 5.10d, h. Instead of the basin-wide spatial scale, further analysis of the co-variation of the annual rainfall from the various grid points of $10^{\circ}\times10^{\circ}$ spatial domain and the selected climate indices of the various regions also showed difference in the magnitude of the correlation coefficient from one

grid point to the next in each region (Figure 5.11). This was indicative of the influence of micro-climate on rainfall variability.

Table 5.1 shows the linkage between rainfall variability and selected climate indices considering various spatio-temporal scales. The magnitudes of the correlation were generally shown to increase as the size of the spatial domain also increased. This showed that the variation in the large-scale ocean-atmosphere interactions could be used to explain the rainfall variability more at a regional- than location-specific spatial scale. Although the correlation would be expected to systematically increase with the reduction in the series temporal resolution (*e.g.* from monthly to annual time scale), it is noticeable that the results tended to vary from one time scale to another. In other words, the drivers of the changes in annual rainfall of the study area may be different from those of other time scales. This implies that investigation of the drivers of rainfall variability should be made in a temporal-scale-specific way. The results of such an investigation based on various temporal scales can be relevant for different water resources applications. For instance, the consideration of point rainfall of high temporal resolution (*e.g.* daily time scale) can be relevant for risk-based applications related to droughts and floods at catchment-scale. Rainfall of seasonal and annual time scales are vital for planning and management of agricultural practices at regional-scale.

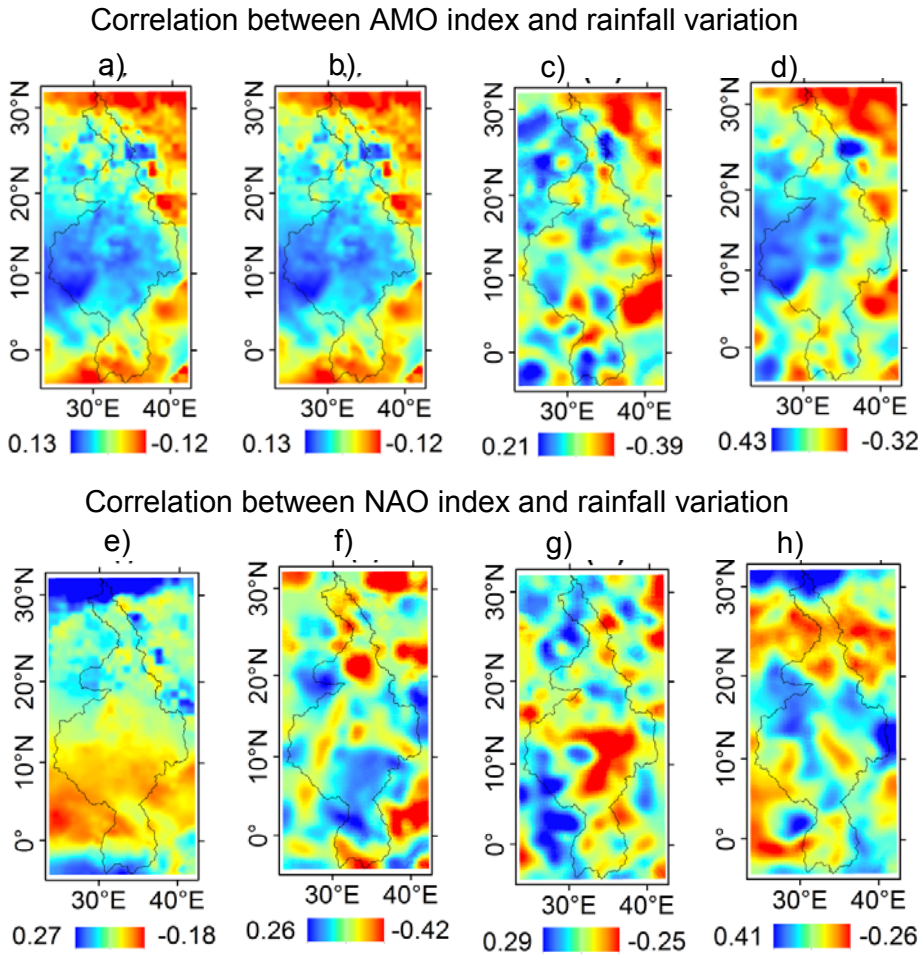


Figure 5.9 Co-occurrence of the variability in rainfall and (a-d) AMO, and (e-h) NAO based on (a, e) monthly, (b, f) JJAS, (c, g) MAM, and (d, h) annual time scales

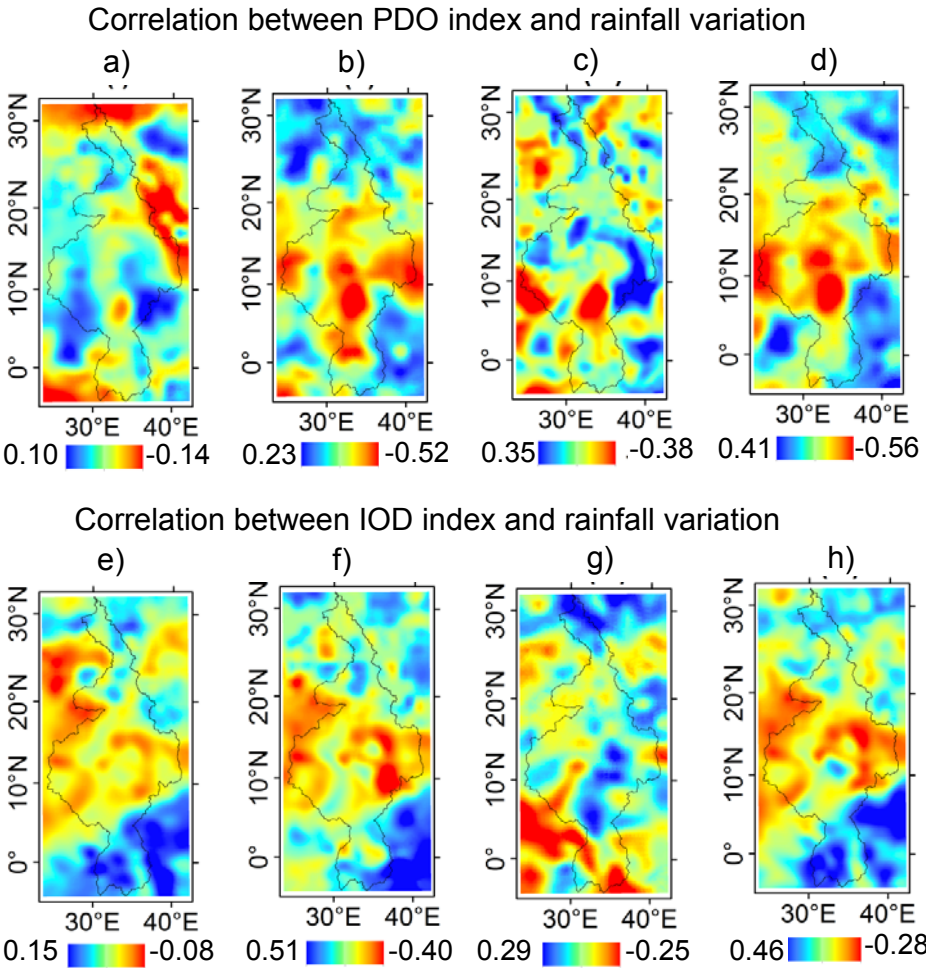


Figure 5.10 Co-occurrence of the variability in rainfall and (a-d) PDO and (e-h) IOD based on (a, e) monthly, (b, f) JJAS, (c, g) MAM, and (d, h) annual time scales

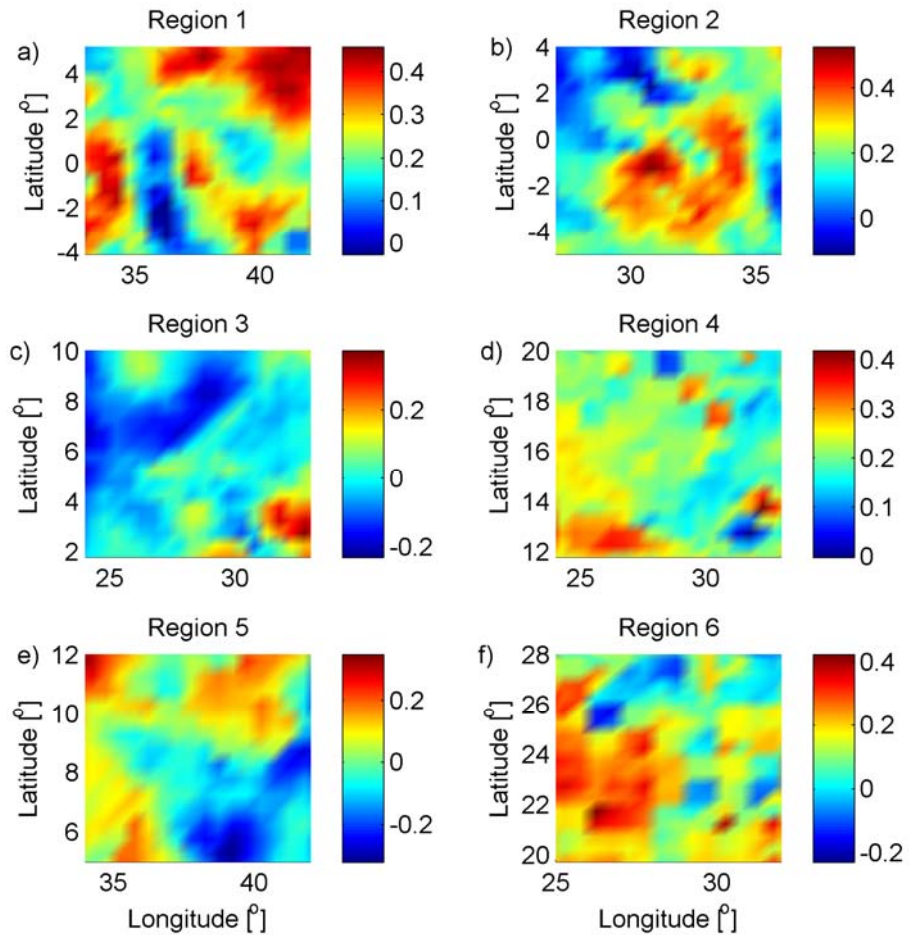


Figure 5.11 Correlation between climate indices and annual rainfall at all the grid points from the 10°x10° spatial domain of region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6; the climate indices IOD, PDO and AMO were used for the correlation values of regions 1 and 2, 3 and 5, and 4 and 6, respectively. The label of horizontal axis for a)-d) is similar to that of e).

5.4 Conclusions and relevance of the findings

5.4.1 Answers to research questions

How does selection of temporal and spatial scales influence results of rainfall variability analyses? Does the choice of the spatio-temporal scales affect the linkage of rainfall variability to large-scale ocean-atmosphere interactions?

The amount of explained variability reduces with increase in the; (i) size of spatial domain, and (ii) temporal resolution of the data. The climate indices were found to explain rainfall variability more suitably at a regional than location-specific spatial scale. The correlation between climate indices and rainfall variability was found to vary from one time scale to another in both magnitude and sign.

5.4.2 Relevance of the findings

These findings are vital in the selection of spatial and temporal scales for relevant information that can be obtained from rainfall variability. To incorporate the effects of climate variability into planning, design and/or management of water resources applications related to floods, high-resolution (such as daily or hourly) data from limited area *e.g.* catchment-based spatial scale can be used. Rainfall of fine temporal scale (*i.e.* characterized by high amount of variability) can be used an input into a hydrological model to predict floods. Planning for applications which depend on rainfall totals, such as agricultural practices, can be done on coarser (*e.g.* regional) than fine spatial scale using information from large-scale ocean-atmosphere interactions based on seasonal or annual-based time scales.

Chapter Six

Flow-rainfall co-variation

This chapter is based on the paper:

Onyutha C, Willems P (2017) Investigation of flow-rainfall co-variation for catchments selected based on the two main sources of River Nile. Stochastic Environmental Research and Risk Assessment, DOI: 10.1007/s00477-017-1397-9

6.1 Introduction

Although the variation of river flow in the Nile basin may be ascribed to the changes in rainfall, a number of studies based on remotely sensed land cover or satellite data (see *e.g.* Elmqvist, 2005; Rientjes *et al.*, 2011) or aerial photographs (Bewket and Sterk, 2005) have reported on the effects of anthropogenic factors on river flow regimes. According to Rientjes *et al.* (2011), forest cover decreased from 50 to 16% in the Gilgel Abay catchment in the Lake Tana basin over the period 1973–2001. Elmqvist (2005) noted that the cropland per household reduced from 0.4 to 0.1 km² over the period 1969–2002 in Sararya Makawi, Sudan. Bewket and Sterk (2005) concluded on an increase in cultivation area in the Chemoga catchment for the period between 1960 and 1999. Flow changes were in these studies linked to the land-use changes. Bewket and Sterk (2005), for instance, related the identified decrease of 0.6 mm/year in the Chemoga catchment flow during the dry season (October to May) between 1960 and 1999 to the increase in cultivation area. The main problem with such flow change studies is that for an accurate analysis, archives are required of aerial photos or satellite images of land cover with high spatial and temporal resolutions and with good quality for long-term periods. Such archives are difficult to obtain for the study area. To partly meet the limitation of such archives, some studies complemented the available land-use and cover data from satellite images with catchment hydrological modeling. The effect of the change in catchment characteristics on the

watershed hydrology can indeed be investigated using hydrological models, by preference to fully distributed process-based models. However, the input data required by such detailed hydrological models are of large amount. Besides, due to their structural complexity and over-parameterization, the parameters of such models are difficult to optimally estimate. Alternatively, conceptual models that are more parsimonious, hence with fewer parameters than the physically-based models, can be applied to assess changes in catchment response in a meteorological-river flow data-based way, but at a lumped catchment scale. Such modeling studies were conducted by Mango *et al.* (2011), and Olang and Fürst (2011) for the Equatorial region; Legesse *et al.* (2003, 2004), Bewket and Sterk (2005), Rientjets *et al.* (2011), and Gebrehiwot *et al.* (2013) for Ethiopia. Based on the land-use scenario investigation using the Soil and Water Assessment Tool (SWAT), Mango *et al.* (2011) concluded for the Mara catchment that the magnitude of the extreme low/high flow would reduce/increase if the conversion of forests to agriculture and grassland in the headwaters of the catchment continued. Olang and Fürst (2011) used the HEC-HMS rainfall-runoff model to investigate the effect of the land-use changes over the period between 1973 and 2000 on the hydrology of the Nyando catchment. The authors found an increase of 16% in the peak discharges over the entire period considered. Using the PRMS model, Legesse *et al.* (2003) found that flow would reduce to about 8% if the dominantly cultivated/grazing land of South Central Ethiopia was to be converted to woodland. Similarly for Lake Abiyata, Legesse *et al.* (2004) noted a remarkable mismatch between the observed and PRMS-modeled lake level over the period 1984–1996 compared with that for 1968–1983. The authors ascribed this discrepancy to human influence on the lake in terms of the direct use of the influent rivers. By dividing the time series over the period 1960–2004 into three parts based on either the political and land management policy changes, Gebrehiwot *et al.* (2013) applied the HBV model to investigate the effect of land-use changes on the runoff flow in the Birr, Upper-Didesa, Gilgel Abbay, and Koga catchments of the Blue Nile basin. According to the authors, although six out of nine parameters of

the HBV model changed significantly over the three periods during the rainfall-runoff modeling, the integrated functioning of the watersheds showed minimal changes.

The problem with the above studies each of which applied only one hydrological or rainfall-runoff model lies in the lack of insight about the influence of model selection on the conclusive flow variation analysis. Moreover, based on the model complexity and set of parameters for calibration, the judgment of the confidence in the selection of a particular model to investigate the effect of land-use change on the flow variation is not a simple task. Other factors such as the change in meteorological conditions need to be addressed as well. Studies by Abtew *et al.* (2009), Camberlin (1997), Taye and Willems (2013), and Tierney *et al.* (2013), gave evidence that the variability in hydro-climatic variables such as rainfall over the study area can be explained by the variation in large-scale ocean-atmosphere interactions.

In this study, because of the data limitation and quality problem for rainfall-runoff modeling in the Nile basin, three rainfall-runoff models *Nedbør-Afstrømnings-Model* (NAM) (Danish Hydraulic Institute DHI, 2007; Madsen, 2000), *Hydrologiska Byråns Vattenavdelning* (HBV) (Bergström, 1976; AghaKouchak and Habib, 2010; AghaKouchak *et al.*, 2013) and *Veralemeend conceptual Hydrologisch Model* (VHM) (Willems, 2014; Willems *et al.*, 2014) were applied. These three models were adopted in this study because they have been recently used by Taye and Willems (2013) (for NAM and VHM), and Gebrehiwot *et al.* (2013) (for HBV) to successfully investigate the effect of meteorological and catchment rainfall-runoff response changes on the flow regimes in the study area. However, in some of these previous studies, manual calibration which is known to be subjective, was adopted. To limit the influence of subjectivity in the model calibration process and address the models' uncertainties, the Generalized Likelihood Uncertainty Estimation (GLUE) of Beven and Binley (1992) was adopted in this study. The model-based findings moreover were

complemented with the analyses of temporal variation in overland flow and trends in the model residuals to support the hypothesis of flow variation driving influence. In explaining the identified trends and temporal variation, special attention was given to the co-variation of flow and rainfall. The final goal was to provide new insights in the spatio-temporal variation of flow along the main River Nile. To do so while taking into account the regional differences in the flow variation from the two main sources of the River Nile, a total of four catchments were considered. The selected catchments included Kyoga and Kagera (from the White Nile region), as well as Blue Nile and Atbara (in the Central region mainly drained by the Blue Nile). As opposed to the previous flow change studies which were mostly limited to catchment scale, the insight from considering the regional differences as in this study is vital for regional planning of water resources of the River Nile.

More specifically, this study aimed at: 1) investigating the co-variation of flow and rainfall, and 2) rainfall-runoff modeling to investigate the evidence of changes in rainfall-flow catchment response behavior. The modeling was done at daily time scale in order to explain aggregated variation at larger temporal scale.

6.2 Selected catchments

For this study, four catchments including Kagera, Kyoga, Blue Nile and Atbara were selected from the Nile basin (Figure 2.7). Brief descriptions of the catchments are given below.

River Kagera catchment located between the Lakes of Victoria, Tanganyika and Kivu has a drainage area of about 60,000 km² and stretches into four countries of East Africa including Burundi, Rwanda, Tanzania, and Uganda. The catchment stretches in the North-South direction from 0°45' to 3°35' S and in the East-West direction from 29°15' to 30°51' E. From the drainage area 184,000 km² of the LVB, River Kagera has the largest catchment area compared to those of other catchments.

The Lake Kyoga which is shallow and situated at the mean elevation of 1,033 m receives flow from the Victoria Nile and the tributaries in the Mount Elgon region. Within the Great Lakes region, Lake Kyoga (though not a great lake itself) acts as an important connection between the two lakes Albert and Victoria with respect to the River Nile flow. Lake Kyoga catchment with the drainage area of about 57,000 km² stretches between 0°20' and 3°40' N (in the North-South direction) and 32°10' and 34°50' E (in the East-West direction). The Eastern boundary of the Lake Kyoga drainage area runs along the Uganda-Kenya border.

The Blue Nile catchment with its outlet situated around the Tuti island between Omdurman and Khartoum North has a drainage area of about 325,000 km². The Blue Nile with a total length of about 1,460 km flows into and out of Lake Tana and emanates from the Ethiopian Highlands based on the two main tributaries including the Dinder and Rahad rivers.

The Atbara catchment which is about 202,650 km² stretches in the East-West direction between 33°45' and 39°45' E and in the North-South direction between 11°30' and 18°55' N. The watershed consists of the 880 km-long River Atbara which is the last tributary of the River Nile. The River Atbara which joins the main Nile about 320 km downstream of Khartoum, has several tributaries a few of which include Tekeze, Bahar El Salam, *etc.*

6.3 Methodology

6.3.1 Co-variation of flow and rainfall

The variation of flow was compared with that of rainfall at both seasonal and annual time scales. In the LVB, two rainy seasons exist *i.e.* MAM and OND (Nicholson, 1996). The JJAS is a dry season. However, in the Northern half (or part) of the Nile basin, JJAS is the main rainy season and the ONDJF is a dry period. Consideration was given to the flow at the catchment outlet and catchment-wide weighted average rainfall based on data at stations upstream of the catchment flow outlet. A zoomed-in analysis such as that for the catchment-scale is important to obtain an insight into how strongly

the micro-climate (micro-scale features) influences the flow-rainfall co-variation. For each selected catchment, comparison was made between the temporal variation of flow at the outlet of the catchment and that of: 1) rainfall at each meteorological station, and 2) catchment-wide average rainfall.

6.3.2 *Rainfall-runoff modeling*

Three rainfall-runoff models including NAM, VHM and HBV were applied under the same meteorological conditions in simulating daily flow at the outlets of two selected main catchments. Kagera catchment at Kyaka Ferry (60,000 km²) in the Equatorial region and the Blue Nile catchment at Khartoum (325,000 km²) from Sudan and Ethiopia were selected based on the two sources of the River Nile as already highlighted in Section 6.2.

The rainfall-runoff modeling was used to assess any possible empirical evidence for the alteration in catchment characteristics in the form of temporal change in catchment runoff response to rainfall, hence between observed and simulated flow. Explanation on this idea can be found in Section 6.3.3. VHM was selected because its model-structure identification and calibration approach nicely fits within the objective of this research to analyze flow and catchment changes in a data-based way. The VHM approach starts from a generalized model structure (Figure 6.1a), which describes the main, lumped catchment-scale rainfall-runoff responses and processes underlying these responses. The model structure includes processes that can be identified from the available meteorological (model input) and river flow (model output) data. This data-based approach also allows identification of changes in catchment-wide responses and related processes. More specifically, Willems (2014) explains that in the VHM approach, the lumped macroscopic catchment responses are analyzed and main processes derived using a step-wise procedure including: 1) separation of observed discharge into overland flow, interflow and base flow, 2) extraction of nearly independent extremes in the form of Peak Over Threshold (POT) events. These time-series processing procedures can be

done using the tool provided by Willems (2009). The different sub-flow components from (1) are used to identify linear or exponential relationships in the various sub-models describing the rainfall fractions contributing to the surface storage, soil moisture storage and groundwater storage. Quick flow is obtained from a combination of the overland and interflow, which are routed separately using a single linear reservoir for each sub-flow component. The baseflow is also routed from the groundwater storage using a linear reservoir. The total runoff is obtained as the sum of the quick flow and base flow.

To take into account possible inconsistency (if any) among the models, the VHM results were, in this study, compared with those from two other, internationally well-established models NAM (DHI, 2007; Madsen, 2000) and HBV (Bergström, 1976; AghaKouchak and Habib, 2010; AghaKouchak *et al.*, 2013) whose structures are presented in (Figure 6.1b) and (Figure 6.1c) respectively. They all consist of a number of inter-related storages for surface, soil and groundwater, relationships describing the influxes in the reservoirs, mainly depending on the time-variable relative soil saturation state (derived from the soil water storage), evapotranspiration losses, and sub-flow routing by means of linear or non-linear reservoir models. Total runoff from either HBV or NAM also comprises the combination of base flow and quick flow.

All the three models use catchment-averaged rainfall and PET as the inputs. The actual evapotranspiration is calculated by the models based on the PET and the soil water storage results. Whereas the original models (especially NAM and HBV) include snow modules, these were left out from the descriptions and application, given the location of our study area in a Tropical region. Whereas catchment areas were fixed apriori, the definitions of the model parameters considered for calibration are presented in Table 6.3.

For each selected catchment, the catchment-wide mean areal rainfall was calculated using the Thiessen polygon method. The PET was computed using the FAO Penman-Monteith method (Allen *et al.*, 1998) considering T_{min} and T_{max} . Apart from the T_{max} and T_{min} , the calculation of PET also requires humidity, radiation and wind speed. For data scarce regions *e.g.* the study area, the procedures for estimating the PET amidst missing data can be found in the FAO Irrigation and Drainage Paper 56. For instance, where wind speed is missing, a value of 2 m/s can be taken as an estimate. This is on the assumption that PET is not highly sensitive to normal ranges of wind speed since the crop height of 0.12 m required for the reference crop is relatively small (Allen *et al.*, 1998). More so, by assuming T_{min} to be low enough as the dew point temperature, the actual vapor pressure can be computed based on the T_{max} and T_{min} after obtaining the maximum and minimum relative humidity, respectively. Solar radiation or the shortwave radiation can be estimated from the duration of bright sunshine. Extraterrestrial radiation can be estimated as a function of latitude, date and time of day.

The calibration (validation) period of the models was 01/01/1965–31/12/1974 (01/01/1975–31/12/2002) for the Blue Nile, and 01/01/1950–31/12/1959 (01/01/1960–31/12/1986) for the Kagera. In other words the calibration (validation) period was 10(28) years for the Blue Nile and 10(27) years for Kagera. Model calibration strategy which also incorporates the uncertainty estimation technique was required for changing the model parameters. Some of such strategies include the GLUE (Beven and Binley, 1992), and Uncertainty Estimation based on Local Error and Clustering (UNEEC) (Solomatine and Shrestha, 2009). The UNEEC method by considering the model residuals to be indicative of the model uncertainty quantifies the uncertainty in three steps including: clustering, model error probability distribution estimation, and building model uncertainty based on the error probability distribution (Solomatine and Shrestha, 2009). As a Bayesian approach, GLUE procedure uses parameters' sets randomized from the prior distribution to infer the output

(posterior) distribution based on the simulations. Because the HBV version used in this study employs the GLUE technique, GLUE was also adopted for NAM and VHM. This was for the uniformity of calibration scheme to obtain an ensemble of simulated flow from each of the models.

Uncertainties from model input series due to observation errors, data limitation, in-filling of missing data values, *etc* lower the accuracy of model predictions. Another source of uncertainty is the incompatibility of the model structure with data availability. Because these uncertainties affect the flow-rainfall relationship, they exacerbate the parameter uncertainty during the calibration of the rainfall-runoff model. The GLUE technique adopted in this study as a framework for the calibration helps to quantify such uncertainties in the form of bounds on the final simulated flow. To implement the GLUE, so many *e.g.* 1000 sets of parameters were randomized based on the stipulated upper and lower limits of parameters. The model was run using each set of the parameters. The optimal set of parameters was taken as that which yielded the highest value of the objective function *e.g.* Nash-Sutcliffe Efficiency (E_{NS}) (Nash and Sutcliffe, 1970). However, for uncertainty quantification, all the sets of parameters and corresponding simulations which satisfied a stipulated objective function were accepted. Normally, the uncertainty in the simulation can be given in terms of the maximum and minimum simulated flow ensemble. For brevity, the uncertainty in the form of 95% CI was stipulated on the flow event corresponding to 10-year return period. The 10-year return period was selected because of its relevance and common use for planning, design and operation of water resource projects. Consider that N_{ACC} denotes the number of the sets of parameters for which objective function was satisfied. For the selected 10-year flow quantile, the simulated flow events from all the N_{ACC} sets of parameters were ranked from the highest to the lowest and the $[0.025 \times N_{ACC}]^{th}$ and $[0.975 \times N_{ACC}]^{th}$ values respectively comprised the upper and lower limits of the 95% CI.

The model performance was evaluated both statistically and graphically. Statistically, the "goodness-of-fit" was evaluated in terms of the E_{NS} . The impact of anthropogenic factors, climate variability and/or change on hydrology seem to increasingly alter the frequency and severity of risk-based water disasters such as floods and dry conditions from their expected normal occurrences. For rainfall-runoff models to be applied for impact investigation, for instance due to climate variability and/or change on extreme flow events, the ability of the models to capture the high and low flow were deemed important for risk-based planning and management. Eventually, graphical plots of Box-Cox (BC) transformed observed *versus* simulated series were made for the maximum as well as minimum flow in each year. The BC (Box and Cox, 1964) transformation (Eq. 6.1) with the parameter (λ) set to 0.25 was applied to give similar weights to the maximum and minimum flow events (f) so as to obtain homoscedastic model residuals. Comparison of the simulated and observed cumulative flow was also made.

$$BC(f) = \frac{f^\lambda - 1}{\lambda} \quad (6.1)$$

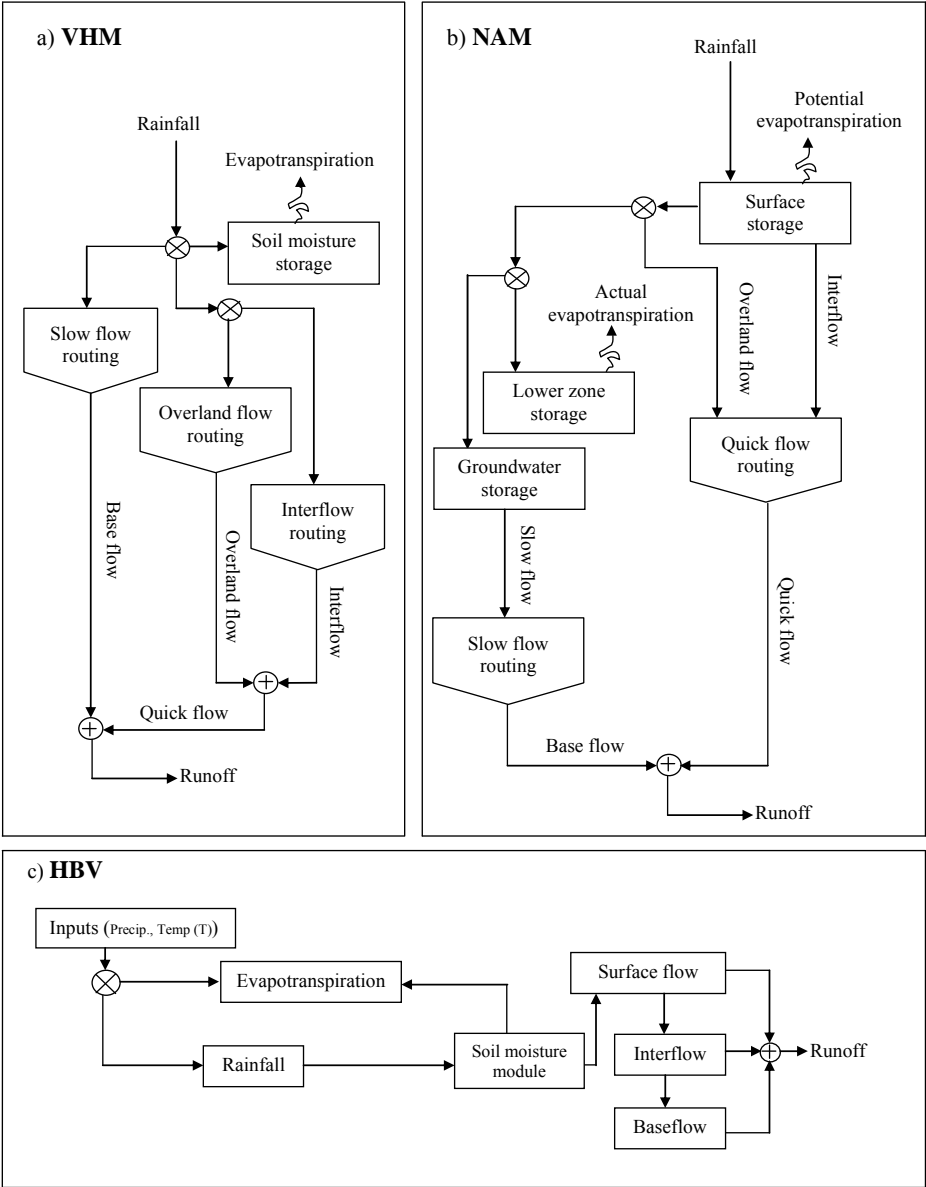


Figure 6.1 General processes of a) VHM, b) NAM, and c) HBV

6.3.3 Analyses of the changes in the flow

According to Merz *et al.* (2012), the flow change linkage to assumed drivers can be done quantitatively in either data- or simulation-based way. Both ways were considered in this study.

a) Data-based approach

A data-based approach was implemented by comparing the correlation between the variation of flow with that of the rainfall series. The rainfall stations from Table 2.4 were used. For a particular station where flow data were from two sources, the series were converted to annual time scale and combined to form longer dataset than that obtained from each source. Finally, common data periods of flow and rainfall including 1940-1986, 1950-2000, 1965-2002 and 1912-1982, for Kagera, Kyoga, Blue Nile and Atbara, respectively were used. High correlation between flow and rainfall was considered to indicate some evidence that the influence of the anthropogenic factors on the catchment runoff generation processes was limited.

b) Simulation-based approach

Because of the non-linearity of the rainfall-runoff relation, the data-based technique was complemented with a simulation-based approach. Three models were applied to study the co-variation of observed flow and catchment rainfall-runoff considering a lumped catchment approach. In case of an unchanging catchment behavior, hence in case of insignificant anthropogenic factors, the temporal flow variation could be assumed to be fully described by the variation in the meteorological model inputs (rainfall and evapotranspiration) after keeping the model parameters constant over time. On the other hand, in case of a change in catchment behavior due to anthropogenic influence, there would be a temporal change in the difference between the observed and modeled runoff flow and sub-flow when model parameters are kept constant. Anthropogenic influences such as deforestation, overgrazing, significant expansion of urbanized areas, *etc* over a given catchment would: 1) affect the amount of infiltration into the soil, 2) alter the amount and velocity of the overland flow, 3) modify the rate and amount of evaporation, *etc*. Hence, these factors would alter the catchment response to the rainfall input. This difference in response should

be visible through the changes in runoff volumes, sub-flow volumes, ratio between sub-flow volumes, model parameters describing the sub-flow response to such times such as the recession constant.

Because of the importance of studying the runoff sub-flow and more specifically the overland flow separately, a numerical digital filter was applied to split the flow into the various sub-components. This discharge splitting was done based on the sub-flow recession constants as applied in the tool provided by Willems (2009).

The simulation-based approach to search for the temporal changes in the overland flow was analyzed in terms of variability using the Quantile Perturbation Method (QPM) (Ntegeka and Willems, 2008; Willems, 2013), cumulative effects of temporal variation (Onyutha, 2016a), and trend analyses based on Mann-Kendall (MK) (Mann, 1945; Kendall, 1975), and Cumulative Sum of rank Difference (CSD) (Onyutha, 2016a; 2016b; 2016c) tests. The QPM was applied directly to the annual maxima, annual minima and annual mean flow. The MK and CSD tests were conducted on the model residuals.

i) Variability analysis using QPM

The QPM uses the given series directly (*i.e.* without rescaling) to obtain quantile anomalies. This allows the QPM outputs to be applicable, for instance, in revising design quantiles to account for the decadal or multi-decadal oscillations or variability in the hydro-meteorological variable. The procedure of applying QPM can be found already described in Section 3.2.1. An elaborate and systematic description of the QPM can be obtained from Ntegeka and Willems (2008) and Willems (2013).

ii) Cumulative effects of temporal variation (Onyutha, 2016a; 2016c)

If for the given series X of size n , another dataset Y is obtained as the replica of X , the rescaled series d in terms of the exceedance and nonexceedance counts of data points can be obtained by (Onyutha, 2016a):

$$d_i = 2 \sum_{j=1}^n \text{sgn}_1(y_j - x_i) - \left(n - \sum_{j=1}^n \text{sgn}_2(y_j - x_i) \right) \text{ for } i = 1, 2, \dots, n \quad (6.2)$$

where,

$$\text{sgn}_1(y_j - x_i) = \begin{cases} 1 & \text{if } (y_j - x_i) > 0 \\ 0 & \text{if } (y_j - x_i) \leq 0 \end{cases} \quad (6.3)$$

$$\text{sgn}_2(y_j - x_i) = \begin{cases} 1 & \text{if } (y_j - x_i) = 0 \\ 0 & \text{if } (y_j - x_i) < 0 \text{ or } (y_j - x_i) > 0 \end{cases} \quad (6.4)$$

The cumulative sum (S_m) of d from Eq. (6.2) is given by (Onyutha, 2016a):

$$S_{m,i} = \sum_{j=1}^i d_j \quad \text{for } 1 \leq i \leq n \quad (6.5)$$

To graphically identify changes in the series, S_m is plotted against time of observations. With reference to the $S_m = 0$ line, positive and negative values characterize various forms of changes, as illustrated in Figure 6.2 for some few simple cases (Onyutha, 2016a; 2016b). Some of the cases that can be identified include: no trend (Figure 6.2a, 1), data with opposite directions of sub-trends over the two halves of the series (Figure 6.2b-c and 2-3), positive trend (Figure 6.2d, 4), negative trend (Figure 6.2e, 5), step downward jump in mean (Figure 6.2f, 6), step upward jump in mean (Figure 6.2g, 7). Although, the S_m (Eq. 6.5) can be used to identify changes

graphically, in fact, it represents pattern of the partial terms that can be statistically used to test the H_0 (no trend) (Onyutha, 2016a; 2016b). This makes the CSD test different from existing methods which purely rely on statistical results.

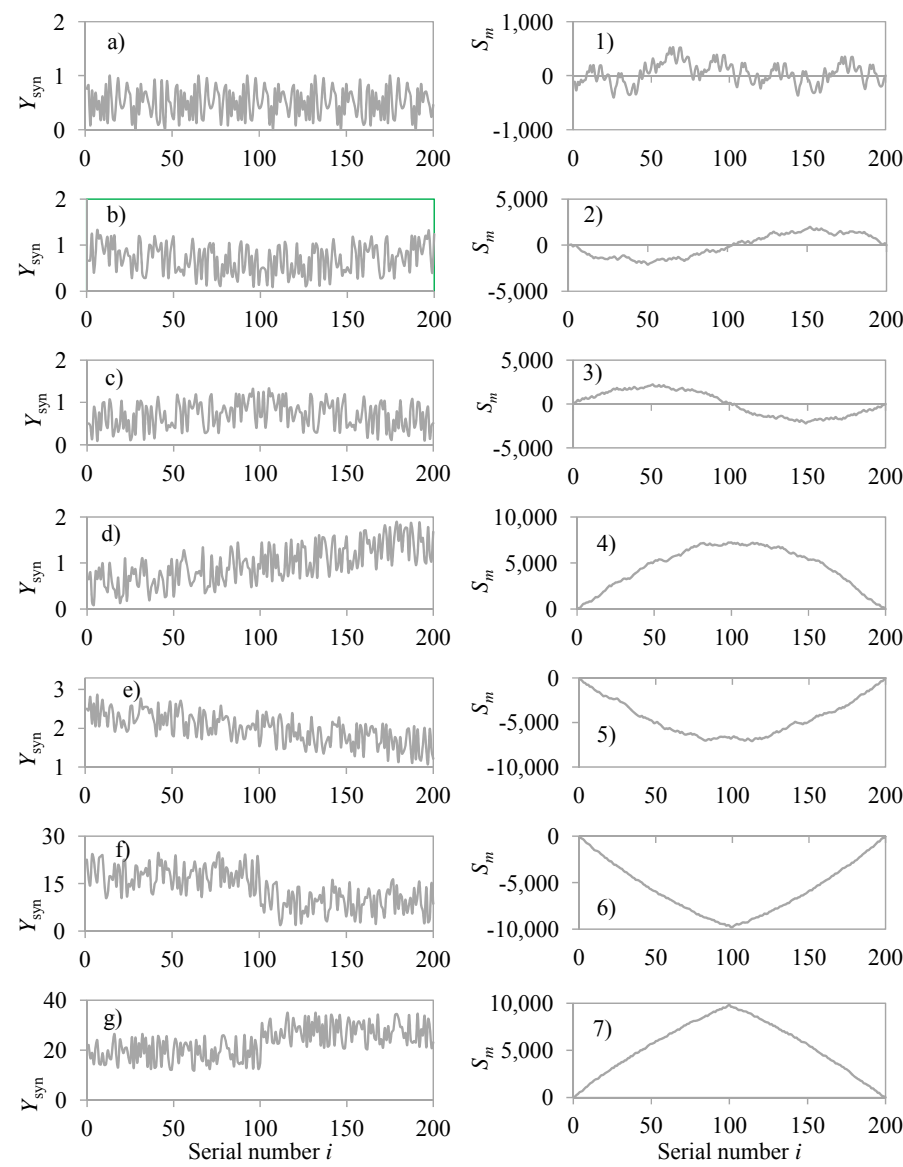


Figure 6.2 Some forms of changes in (a-g) synthetic series Y_{syn} and (1-7) corresponding cumulative effects of the temporal variation

iii) Analyses of trends in model residuals

In the simulation-based procedure, with the premise that the models fully capture the catchment behavior, any deviation between observed and simulated flow can be linked to internal disturbances which may include forest cover change, urbanization, river engineering, dam construction, *etc* (Harrigan *et al.*, 2014). In this study, any persistent deviation between the observed and modeled flow was deemed to be reflected in the rejection of H_0 (no trend in model residuals). According to Onyutha (2016c), the uncertainty in trend analyses due to the influence from the selection of a particular method of change detection is not negligible for series with persistent fluctuations. Eventually, the H_0 (no trend in model residuals) was tested using two (*i.e.* MK and CSD) trend tests.

The Mann-Kendall (MK) (Mann, 1945; Kendall, 1975) test

For the MK test, scores were assigned to the model residuals. The MK (Mann, 1945; Kendall, 1975) test statistic S is defined as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (6.6)$$

where x_j and x_i are the sequential data values in a sample of size n , and

$$\text{sgn}(x_j - x_i) = \begin{cases} 1 & \text{if } (x_j - x_i) > 0 \\ 0 & \text{if } (x_j - x_i) = 0 \\ -1 & \text{if } (x_j - x_i) < 0 \end{cases} \quad (6.7)$$

For $n \geq 8$, S is approximately normally distributed with the mean $E(S) = 0$ and variance $V(S)$ given by (Mann, 1945; Kendall, 1975):

$$V(S) = \frac{1}{18} n(n-1)(2n+5) \quad (6.8)$$

When tied data points exist, $V(S)$ becomes:

$$V(S) = \frac{1}{18} \left(n(n-1)(2n+5) - \sum_{k=1}^h g_k(k-1)(2k+5) \right) \quad (6.9)$$

where h is the number of tied groups, and g_k is the number of observations in the k^{th} group.

The standardized MK test statistic Z_{MK} which follows the standard normal distribution with mean (variance) of zero (one) is given by:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{for } S < 0 \end{cases} \quad (6.10)$$

Positive and negative values of S indicate increasing and decreasing trends respectively. The influence of auto-correlation on the variance of S can be corrected based on the procedure suggested by Yue and Wang (2004). The H_0 (no trend) is rejected if $|Z_{MK}|$ is greater than or equal to the standard normal variate $Z_{\alpha_s/2}$ where $\alpha_s\%$ is the significance level; otherwise the alternative hypothesis H_1 (there is trend) is not accepted.

The CSD test (Onyutha, 2016a; 2016b; 2016c)

The CSD trend statistic T_{CSD} is computed using (Onyutha, 2016a):

$$T_{CSD} = \frac{6}{(n^3 - n)} \sum_{i=1}^{n-1} S_{m,i} \quad (6.11)$$

where S_m is from Eq. (6.5). An upward/downward monotonic trend is indicated by a positive/negative value of T_{CSD} . The distribution of T_{CSD} is

approximately normal with the mean of zero and variance (V_1) given by (Onyutha, 2016b; 2016c):

$$V_1 = \frac{1}{n-1} \left(1 - \frac{10}{17} c^2 - \frac{7}{17} c \right) \quad (6.12)$$

where c (Eq. 6.12) is the measure of ties in the data such that

$$c = \frac{-1}{n^2 - n} \left(n - \sum_{i=1}^n \sum_{j=1}^n \text{sgn}_2(y_j - x_i) \right) \quad (6.13)$$

and $\text{sgn}_2(y_j - x_i)$ is as defined in Eq. (6.4).

The standardized statistic of the CSD test Z_{CSD} which follows the standard normal distribution with mean (variance) of zero (one) is given by Eq. (6.14). At the significance level $\alpha_s\%$, the H_0 (no trend) is rejected if the p -value (probability value, p) based on $|Z|$ is less than the nominal α_s ; otherwise, the H_1 (there is trend) is not accepted. To account for the effect of persistent fluctuations on V_1 , Z_{CSD} can, generally, be computed using

$$Z_{CSD} = \frac{T_{CSD}}{\sqrt{V_2}} \quad (6.14)$$

where, according to (Onyutha, 2016b),

$$V_2 = V_1 \times \left| 1 + \frac{2}{n(n^2 - 3)} \times \sum_{k=1}^{n-2} (n-k)^3 r_k^\alpha \right| \quad (6.15)$$

and, r_k^α is the *lag-k* serial coefficient (significant at $\alpha_s\%$ level) and is computed after detrending the original series. The CSD trend test is implemented in the tool CSD-NAIM which can be freely downloaded online (see link <http://dx.doi.org/10.13140/RG.2.1.3691.7362/1> accessed: 01.09.2017).

6.4 Results and discussion

6.4.1 Co-variation of flow and rainfall

Figure 6.3 shows the temporal variation of the mean annual flow at the catchment outlet and the catchment-averaged rainfall for common data periods as described in Section 6.3.3. The closeness between the flow and rainfall variation is visibly evident. Further still, the H_0 (no correlation between rainfall and flow) at the significance level α_s of 5% was rejected for all the selected catchments, *i.e.* Kagera and Kyoga, Blue Nile and Atbara (Figure 6.3a–d).

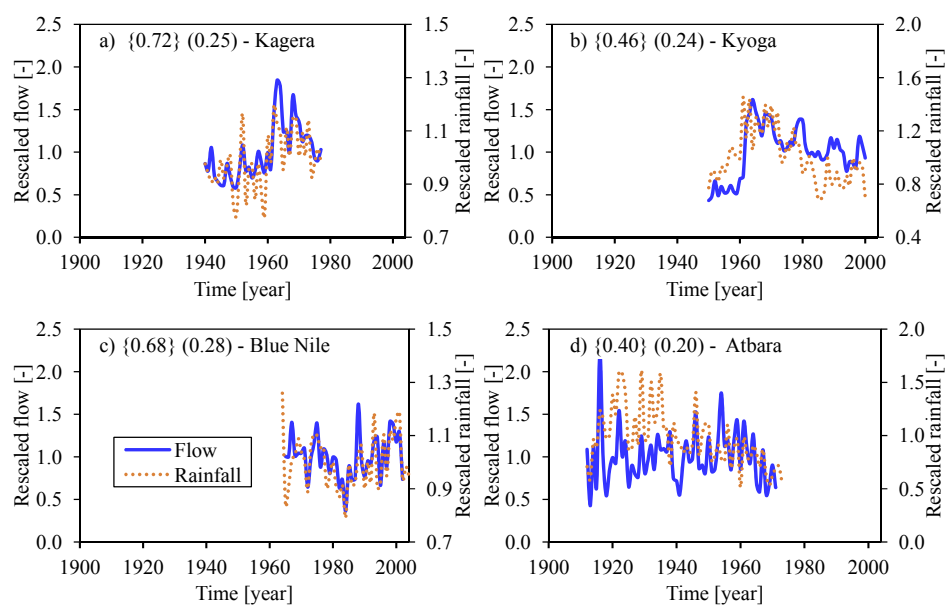


Figure 6.3 Variation in the mean annual flow and catchment-wide rainfall in a) Kagera, b) Kyoga, c) Blue Nile, and d) Atbara; the label in “{ }” shows the correlation between the rainfall and flow, while “()” comprises the correlation critical value at the significance level α_s of 5%; and the legends for charts a)–b) and d) are similar to that of c)

Table 6.1 shows the correlation between the variation of mean seasonal and mean annual flow at the catchment outlet and the rainfall from

individual stations upstream of the outlet. Although there was some insignificant anti-correlation especially in the MAM (JJAS) season for Blue Nile and Atbara (Kagera and Kyoga), generally the correlation for the annual time scale was significantly positive for all the selected catchments. This realization was also obtained for the seasons from which the variation in the annual rainfall emanates, *i.e.* OND (for Kagera and Kyoga) and JJAS (for Blue Nile and Atbara). Conclusively, even without considering other factors such as the changes in land cover or land-use, the variability in rainfall could adequately explain the variation in seasonal and annual flow over the selected data period.

Table 6.1 Correlation between the variation of rainfall and river flow

| Rainfall | MAM | JJAS | ONDJF | Annual | Rainfall | MAM | JJAS | OND | Annual |
|------------------|-------|-------------|-------------|-------------|---|-------------|-------|-------------|-------------|
| Blue Nile {0.28} | | | | | Kyoga {0.24} | | | | |
| Blu1 | -0.04 | 0.29 | 0.62 | 0.34 | Kyo1 | 0.28 | -0.12 | 0.27 | 0.32 |
| Blu2 | -0.11 | 0.34 | 0.57 | 0.32 | Kyo2 | 0.25 | 0.04 | 0.26 | 0.41 |
| Blu3 | -0.21 | 0.47 | 0.53 | 0.5 | Kyo3 | 0.23 | 0.13 | 0.36 | 0.36 |
| Blu4 | 0.1 | 0.46 | 0.09 | 0.31 | Kyo4 | 0.21 | 0.12 | 0.36 | 0.4 |
| Blu5 | -0.15 | 0.69 | 0.26 | 0.55 | CW | 0.32 | 0.06 | 0.38 | 0.46 |
| CW | -0.12 | 0.69 | 0.59 | 0.68 | Kagera {0.28} | | | | |
| Atbara | | | | | Kag1 | 0.38 | 0.12 | 0.36 | 0.64 |
| Abt1 | -0.06 | 0.37 | -0.06 | 0.36 | Kag2 | 0.11 | 0.24 | 0.52 | 0.54 |
| Abt2 | -0.16 | 0.44 | 0.06 | 0.44 | Kag3 | 0.06 | -0.14 | 0.34 | 0.29 |
| Abt3 | -0.09 | 0.59 | -0.08 | 0.58 | Kag4 | 0.37 | 0.21 | 0.39 | 0.53 |
| Abt4 | -0.24 | 0.57 | 0.01 | 0.61 | CW | 0.34 | 0.18 | 0.6 | 0.72 |
| Abt5 | -0.02 | 0.36 | 0.11 | 0.35 | For bold values H_0 (no correlation) was rejected at the significance level of 5%. CW stands for Catchment-Wide. | | | | |
| Abt6 | 0.12 | 0.42 | -0.03 | 0.38 | | | | | |
| CW | -0.04 | 0.42 | 0.05 | 0.4 | | | | | |

The value in "{ }" denotes the correlation critical value at the significance level of 5%.
The critical values for Atbara catchment were 0.22 (Abt1), 0.37 (Abt2), 0.35 (Abt3), 0.37 (Abt4), 0.22 (Abt5), 0.35 (Abt6), and 0.22 (CW).

6.4.2 Rainfall-runoff modeling

Figure 6.4 shows the time series of the observed and simulated Blue Nile and Kagera flow. The model parameters considered for calibration using the GLUE technique can be found listed in Table 6.3. Though for brevity the uncertainty bounds based on GLUE for the entire calibration period were not included in Figure 6.4, the difference between the upper and lower

limits of the 95% CI on the simulated 10-year flow event as a percentage of the 10-year empirical flow quantile for HBV, NAM and VHM was respectively 35.8, 29.2, 40.3% for Blue Nile, and 41.2, 51.8, 46.2% for Kagera. Generally, the predictive uncertainties were higher for Kagera than those of the Blue Nile. Furthermore, this observation is supported by the visual judgment considering each model in which again better modeled results were obtained for the Blue Nile (Figure 6.4a–b) than for the Kagera (Figure 6.4c–d) catchment. This realization was because of the larger catchment size and stronger seasonal differences for the Blue Nile which led to smoother catchment response to rainfall than that for the Kagera.

Table 6.2 shows the statistical assessment of the models' performance. The values of the E_{NS} for calibration were acceptably high. Consistent with the visual judgment from Figure 6.4b and d, the model performance (though reasonably good) in the validation was better for the Blue Nile than that of the Kagera catchment. In reference to the model calibration results, the validation E_{NS} for VHM, HBV, and NAM exhibited a drop by 8, 10 and 8% (for the Blue Nile) and 13, 14 and 12% (for Kagera) respectively. In order to check whether this drop in the E_{NS} for Kagera could be explained by model over-calibration, the validation input series were considered as independent model inputs. When calibration based on the GLUE scheme was repeated for this new input, there was no significant increase in the E_{NS} values compared to those for the validation shown in Table 6.2. One reason for the E_{NS} drop could be the lower data quality taking into account that some measurement stations tend to be irregularly operational due to poor maintenance. Another potential reason could be the difficulty in the models to capture the dynamics of runoff generation at the catchment level especially due to the recovery of the Equatorial region from the step jump in the flow mean which occurred at the beginning of the validation period for Kagera (*i.e.* 1960–1986). Because the E_{NS} values for all the models were higher than 0.6 for all the validation runs, the modeled flow was deemed adequate for the empirical investigation of the possible change in catchment rainfall-runoff response.

Table 6.2 Statistical evaluation of model performance

| | Blue Nile | Kagera | Blue Nile | Kagera |
|--------|-------------|-----------|------------|-----------|
| | Calibration | | Validation | |
| Period | 1965-1974 | 1950-1959 | 1975-2002 | 1960-1986 |
| VHM | 0.76 | 0.74 | 0.70 | 0.65 |
| HBV | 0.77 | 0.73 | 0.69 | 0.63 |
| NAM | 0.78 | 0.70 | 0.71 | 0.62 |

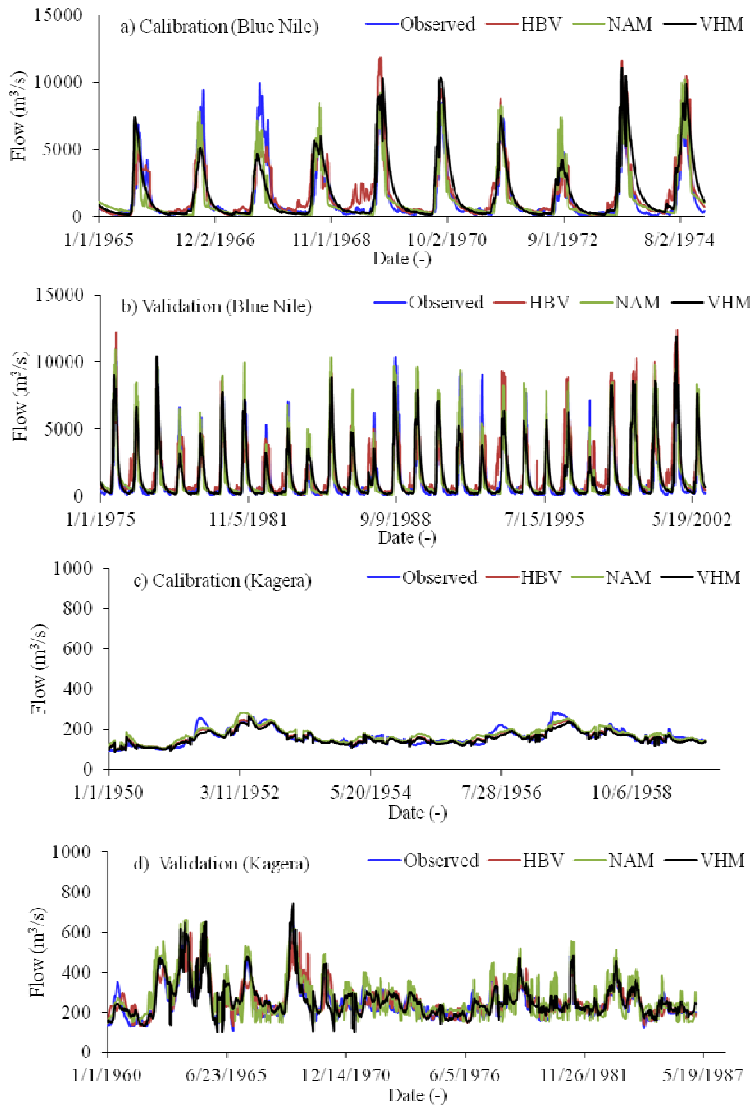


Figure 6.4 Simulated *versus* observed daily flow for a)–b) Blue Nile, and c)–d) Kagera catchments

Figure 6.5 shows the graphical assessment of the "goodness-of-fit" between the Box-Cox transformed observed and modeled daily maximum and minimum flow in a year. Good model performance was obtained for both the annual maximum (Figure 6.5d) and minimum (Figure 6.5e) flow of Kagera catchment. For the Blue Nile, though the performances of all the models were good for the annual maxima (Figure 6.5a), the HBV model tended to over-estimate most of the annual minima (Figure 6.5b). With respect to the cumulative flow volumes (Figure 6.5c and f), the models gave slight over-estimations. For Kagera catchment, the slight over-estimations might have been due to the lower performance of the simulation results from the validation period. Note that because the Blue Nile flow has stronger seasonal variation, the flow scatter was more wide-spread for the Blue Nile (Figure 6.5a–b) than for the Kagera (Figure 6.5c–d). Based on these performance evaluation results, it was generally concluded that the three models were jointly suitable for making an assessment of potential temporal changes in the catchment rainfall-runoff responses.

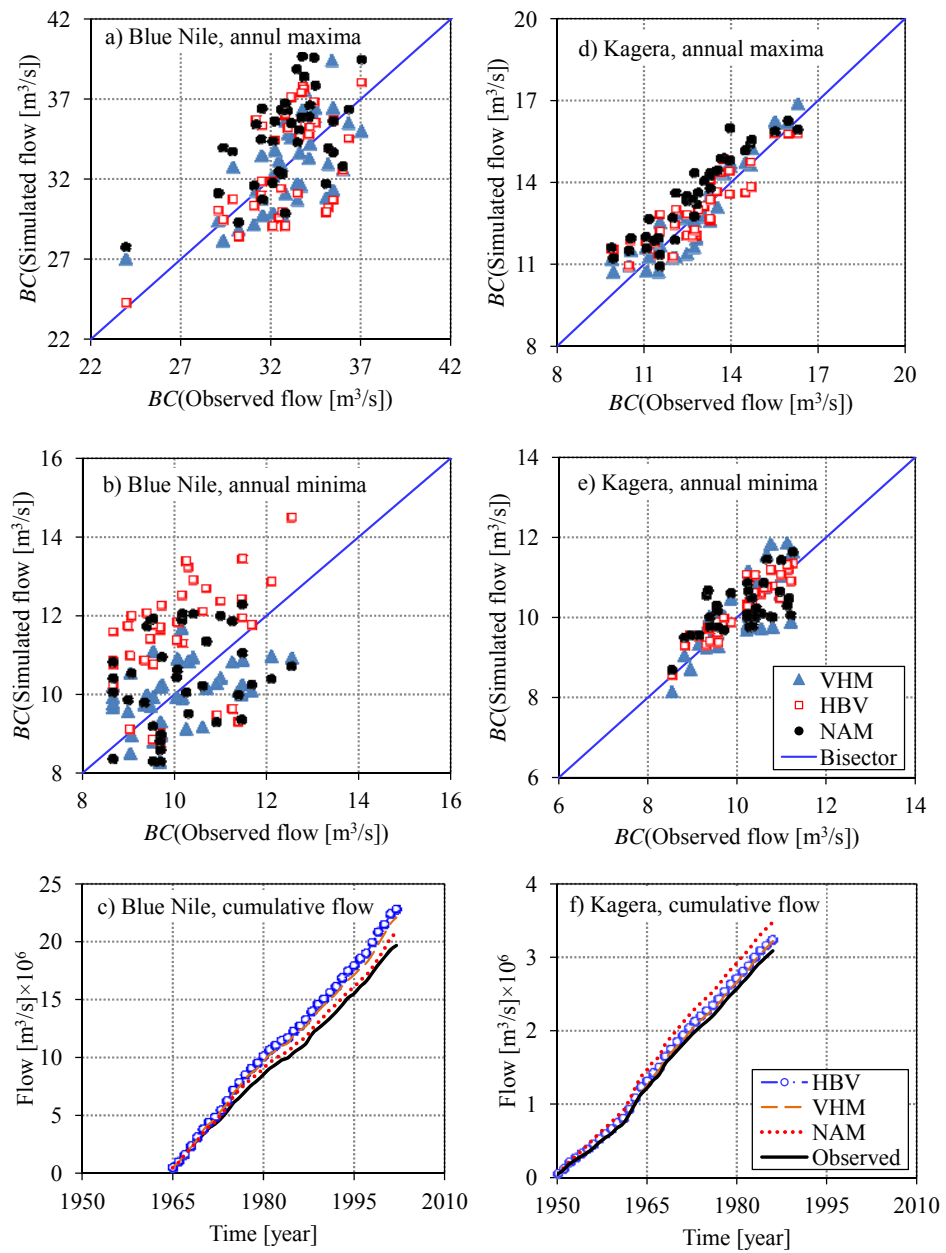


Figure 6.5 Evaluation of the model performance considering Box-Cox (*BC*) transformed (a, d) maximum flow in a year, (b, e) minimum flow in a year, and (c, f) cumulative flow; the legends of charts (a-b, d) and c) are the same as those in e) and f) respectively

Table 6.3 List of model parameters considered for calibration

| Model and list of parameters | | Blue Nile | Kagera |
|---|----------------------------|-----------|--------|
| VHM | | | |
| <i>Soil moisture (storage) sub-model</i> | | | |
| Maximum soil water content: | U_{\max} (mm) | 1950 | 2420 |
| Soil water content at maximum evapotranspiration | U_{evap} (mm) | 300 | 964 |
| Initial soil water content: | U_{init} (mm) | 650 | 20 |
| Model coefficients: | c_1 | 1 | 0.9 |
| | c_2 | 0.67 | 0.1 |
| <i>Overland flow sub-model</i> | | | |
| Model coefficients: | c_1 | -11.07 | -7.09 |
| | c_2 | -0.35 | 2.93 |
| | c_3 | 0.55 | 0.99 |
| | c_4 | 3.12 | 0.82 |
| <i>Interflow sub-model</i> | | | |
| Model coefficients: | c_1 | -8.99 | -9.3 |
| | c_2 | -1.33 | 0.43 |
| | c_3 | -0.15 | 0.65 |
| | c_4 | 3.12 | 3.32 |
| <i>Flow routing recession constant</i> | | | |
| Base flow: | K_B (day) | 120 | 50 |
| Interflow: | K_I (day) | 40 | 40 |
| Overland flow: | K_O (day) | 20 | 20 |
| HBV | | | |
| Shape coefficient | β (-) | 4.32 | 1.04 |
| Field capacity | F_C (mm) | 171.52 | 1058.7 |
| Soil permanent wilting point | P_{wp} (mm) | 97.63 | 130.08 |
| Near surface flow storage coefficient | K_0 (day ⁻¹) | 0.08 | 0.06 |
| Threshold water level | T_{WL} (mm) | 2.23 | 2.27 |
| Interflow storage coefficient | K_1 (day ⁻¹) | 0.03 | 0.05 |
| Baseflow storage coefficient | K_2 (day ⁻¹) | 0.02 | 0.02 |
| Percolation storage coefficient | K_p (day ⁻¹) | 0.04 | 0.11 |
| NAM | | | |
| Maximum soil water content in the surface storage | U_{\max} (mm) | 40 | 0.72 |
| Maximum soil water content in lower zone storage | L_{\max} (mm) | 250 | 0.25 |
| Time constant for overland flow | CQ_{OF} (day) | 0.3 | 0.01 |
| Time constant for interflow | CK_{IF} (day) | 600 | 20 |
| Time constant for overland and interflow routing | $CK_{1,2}$ (day) | 34 | 400 |
| Time constant for baseflow | CK_{BF} (day) | 4000 | 3980 |
| Threshold for overland flow | T_{OF} | 0.23 | 0.2 |
| Threshold for interflow | T_{IF} | 0.1 | 0.5 |
| Threshold for groundwater recharge | T_G | 0.2 | 0.4 |

Figure 6.6 shows changes in observed and simulated daily maximum and minimum overland flow in each year in terms of the QPM quantile anomalies for moving block lengths of 15 years. To interpret the results in Figure 6.6, an anomaly of, say, 5% for a given year indicated that the extreme quantiles in the sub-period centered on this year were on average 5% higher than the quantiles obtained based on the full time series (taken as the reference *i.e.* anomaly of zero). For the annual maximum and mean overland flow at the Blue Nile (Figure 6.6a, c), there were oscillation highs (OHs) in the mid-1960s to late 1970s, and in the 1990s. There was an oscillation low (OL) from the late 1970s to around 1990. Taye and Willems (2013) found similar results especially for the annual maximum flow in the upper Blue Nile catchment. However, for the annual minimum flow (Figure 6.6b), the period from the mid-1960s to late 1970s (early 1980s to late 1990s) was characterized by an OH (OL). For Kagera, the period 1950-1959 (1960-1986) was characterized by an OL (OH) (Figure 6.6d-f). On a side note, some slight over- or under-estimations of the observed changes were demonstrated in the QPM anomalies.

Figure 6.7 shows graphical results of changes in the observed and modeled overland flow. It can be seen that negative (positive) sub-trend occurred in annual mean and maximum flow of Blue Nile (Figure 6.7a, c) over the period 1965-1983 (1984-2000). However, the annual minimum flow was characterized by a negative trend (Figure 6.7b). For Kagera, the step jump in mean of the flow around 1960 is noticeable for the annual mean, maximum and minimum flow (Figure 6.7d-f). Despite some slight under- or over-estimation, the cumulative effects of temporal variation in observed overland flow were, generally, well reproduced by the model simulations.

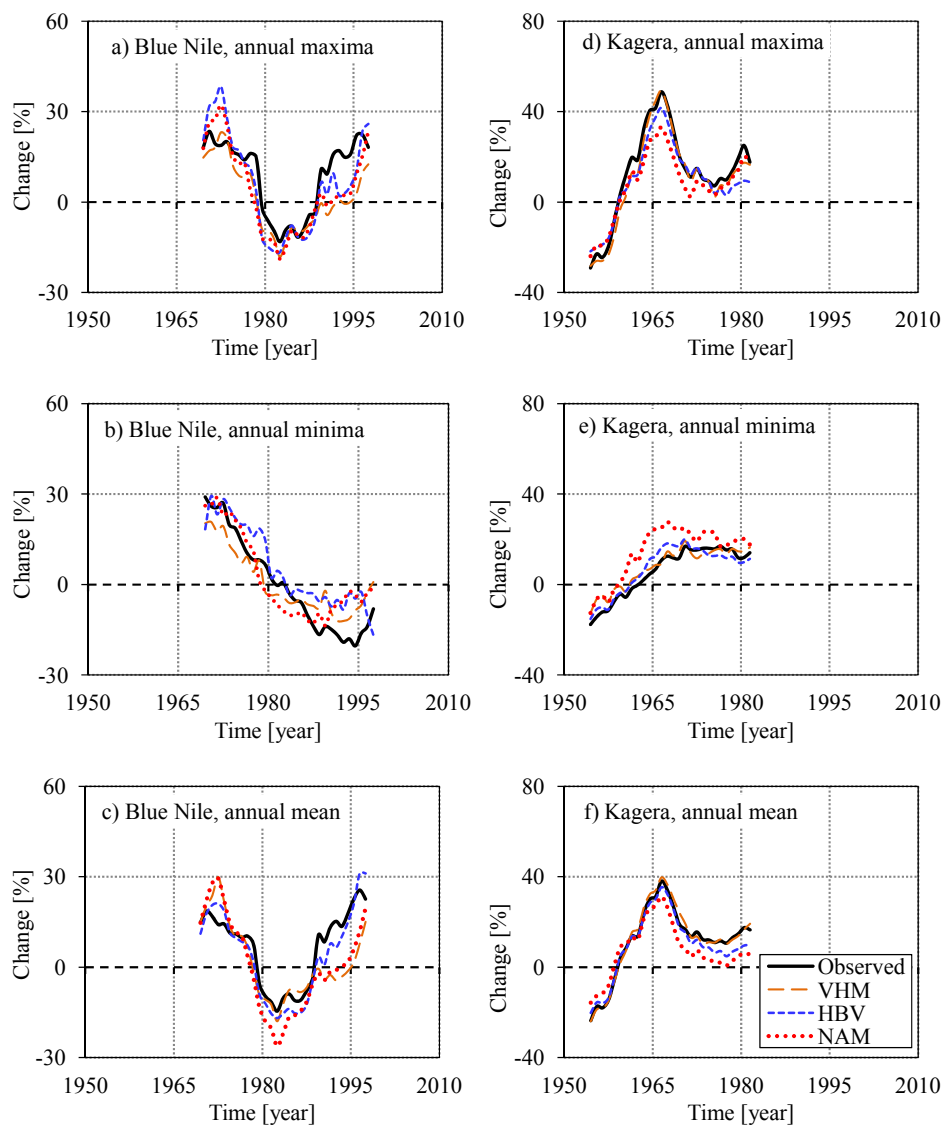


Figure 6.6 QPM results for observed and modeled overland flow in terms of anomalies in the (a, d) annual maxima, (b, e) annual minima and (c, f) annual mean; the legends of charts (a-e) are similar to that in f)

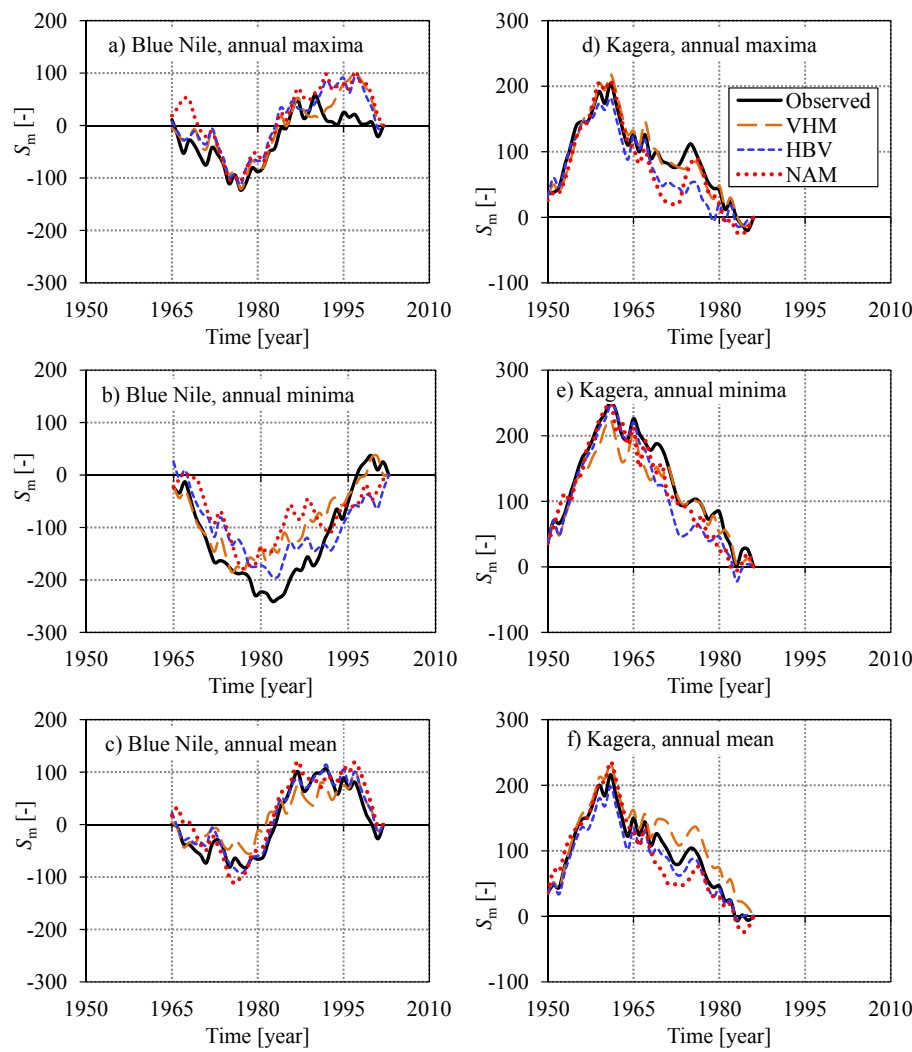


Figure 6.7 Cumulative sum of variation in observed and modeled overland flow in terms of the (a, d) annual maxima, (b, e) annual minima and (c, f) annual mean; the legends of charts (a-c and e-f) are similar to that in d)

Table 6.4 shows the probability values of the trend tests conducted on model residuals computed based on the annual maxima, annual minima and annual mean flow. For the Blue Nile catchment, the p -values were greater than the nominal significance level α_s of 0.05 for all the models. This showed that the magnitude of the model residuals did not depend on the time of observations. In other words, the H_0 (no trend in residuals) was not

statistically rejected at the significance level α_s of 5%. However for Kagera, H_0 was rejected at 5% level based on the MK test at the level at 5% level for the residual of HBV computed using annual minimum flow. Although this might suggest some possible change in catchment behavior, it was not deemed conclusive since the H_0 was not rejected at 5% level based on the results from the two other models *i.e.* NAM and VHM. Moreover for the CSD test, the H_0 was not rejected at 5% level for the same HBV residuals. Generally, the p -values based on the residual from the annual minima flow were lower than those from the annual mean or maxima. Of course, this was not surprising because the structures of most hydrological models are traditionally designed to capture rather the flooding events than low flow conditions (Staudinger *et al.*, 2011). Though not shown in Table 6.4, the residual trend directions in terms of their signs were found different among the models in some cases (see Figure 6.8). Residual trends were in the same direction for the annual maxima and mean flow of the Blue Nile basin, and the annual minima flow of Kagera (Figure 6.8a, c, e). However, trend in the residual was negative (positive) for VHM (HBV and NAM) for residuals from Kagera annual maxima (Figure 6.8d), and VHM and NAM (HBV) for residuals of the annual minima of Blue Nile and annual mean of Kagera (Figure 6.8b, f). The dissimilarities in the trend directions could have been because of the difference between the models in terms of their structures and sets of calibration parameters used to capture the runoff generation dynamics.

Table 6.4 The p -values of MK and CSD trend tests

| Residual based on : | Model | Blue Nile | | Kagera | |
|---|-------|-----------|----------|--------------|----------|
| | | MK test | CSD test | MK test | CSD test |
| Annual mean | HBV | 0.741 | 0.593 | 0.628 | 0.717 |
| | NAM | 0.851 | 0.845 | 0.890 | 0.503 |
| | VHM | 0.735 | 0.821 | 0.700 | 0.854 |
| Annual maxima | HBV | 0.991 | 0.841 | 0.927 | 0.779 |
| | NAM | 0.371 | 0.418 | 0.238 | 0.164 |
| | VHM | 0.867 | 0.853 | 0.953 | 0.876 |
| Annual minima | HBV | 0.404 | 0.513 | 0.039 | 0.062 |
| | NAM | 0.428 | 0.402 | 0.114 | 0.211 |
| | VHM | 0.257 | 0.325 | 0.208 | 0.301 |
| For bold value, H_0 (no trend) was rejected at the significance level of 5% | | | | | |

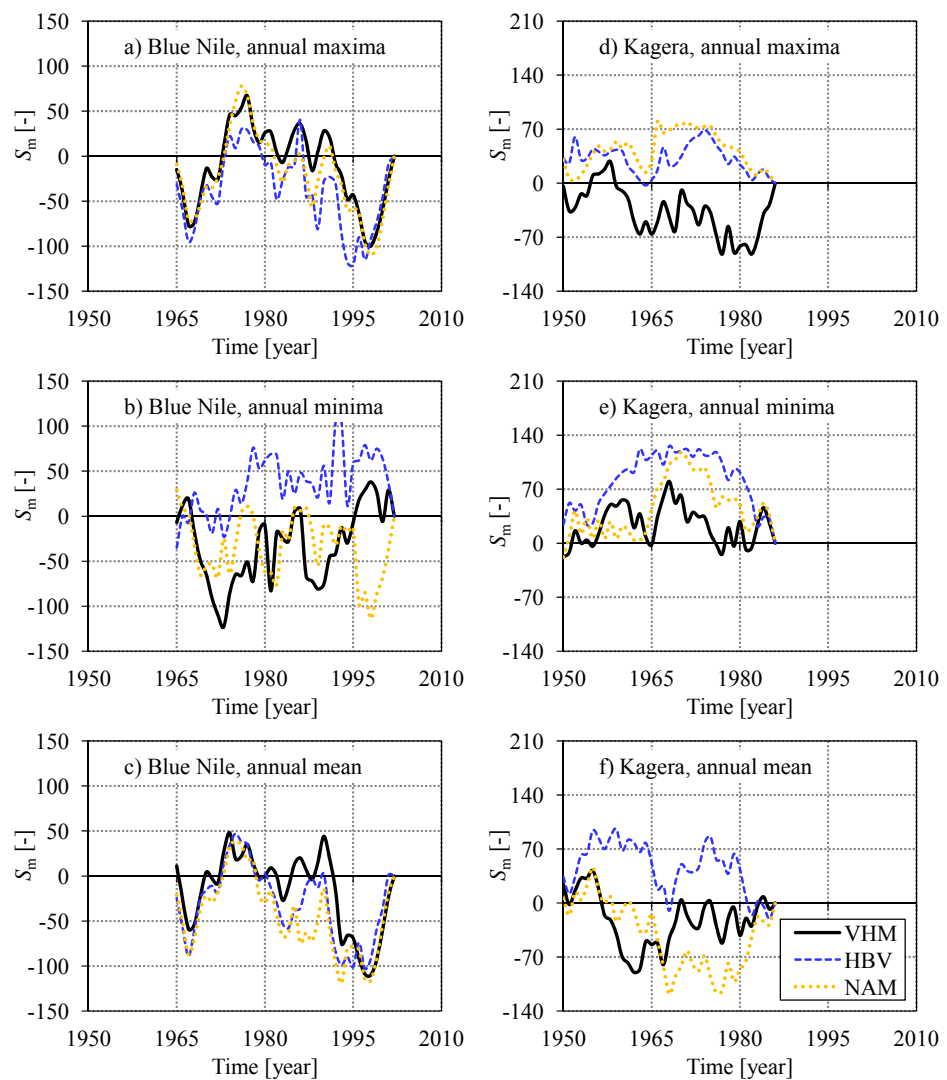


Figure 6.8 Cumulative sum of variation in model residuals from (a, d) annual maxima, (b, e) annual minima and (c, f) annual mean; the legends of charts (a-e) are similar to that in f)

It is worth mentioning that some minor discrepancies between the observed and modeled flow were demonstrated in the QPM results as well as the CSD-based graphical diagnoses of changes. These discrepancies could be indicative of some possible slight changes in the catchment behavior, which may be linked to an anthropogenic influence. A number of

previous studies also ascribed hydrological changes in the different parts of the Nile basin to anthropogenic factors including: human influence on the Lake Tana in terms of abstractions from the influent rivers (Legesse *et al.*, 2004), an increase in cultivation area in the Chemoga catchment (Bewket and Sterk, 2005), deforestation for agriculture in the Gilgel Abay catchment (Rientjes *et al.*, 2011), increase in the population and intensification of agriculture in Sudan (Elmqvist, 2005), conversion of forests to farming land in the Mara catchment (Mango *et al.*, 2011), *etc.*

Unlike in some other studies, the significant catchment changes and their effects on catchment runoff flow were not confirmed in this study. Despite the slight over- or under-estimation of the observed changes for both selected catchments, the outputs from all the models indeed exhibited close agreement in capturing the flow variability or quantile changes (by the QPM). Moreover, the H_0 (no trends in the model residuals) was generally not rejected at the significance level α_s of 5% level based on both the MK and CSD test. These results from the QPM for variability, and MK and CSD tests for trend analyses generally indicated that the changes in the catchment behavior in transforming the rainfall input into runoff were minimal over the data period considered in this study. In related studies, similar conclusions were made by *e.g.* Taye and Willems (2013) and Gebrehiwot *et al.* (2013), all done for the period 1960–2004. According to Gebrehiwot *et al.* (2013), the integrated functioning of the catchments including Birr, Upper-Didesa, Gilgel Abbay, and Koga sub-catchments of the Blue Nile as a main source of the River Nile showed minimal changes. Taye and Willems (2013) also concluded that there were no discernible changes noted for the Blue Nile catchment response. Therefore as shown in this study, the temporal flow variation in the catchments selected from the two main sources of the River Nile may be primarily linked to the rainfall variation.

According to Harrigan *et al.* (2014), despite the statistical detection of trends, rigorous attribution is required in decision making on long-term

management and adaptation strategies. In the same line, the call for increased rigor in attribution through consistency, inconsistency and provision of confidence statement was made by Merz *et al.* (2012). Furthermore, Harrigan *et al.* (2014) suggested the Method of Multiple Working Hypothesis (MMWHs) as a systematic examination of known drivers to explain the full signal of change. In line with the MMWHs, some of the working hypotheses (which were not investigated in this study) but deemed to potentially influence the catchment behavior (though insignificantly) across the study area and should be considered for further research in a combined way include: urbanization, forest cover transition, water abstractions or diversions, agricultural land-use and management change, *etc.* Another factor which could not be ruled out in influencing the change detection was the questionable quality of hydro-meteorological data in the study area. Once the large data requirement for attribution become manageable in the future, an interesting attempt would be to expose the synergistic interaction (if any) of the drivers of the flow changes in the various catchments of the Nile basin.

Some of the possible limitations in applying the methodology used in this study are as follow:

- a) It may be possible that the mismatch between the observed and modeled flow used to give insight about the change in catchment behavior can be due to the questionable data quality. Some of the possible data-related uncertainties could be the observation errors, in-filling of missing values, errors in derivation of the hydro-climatic variables *e.g.* rainfall, *etc.*
- b) In case the changes in the catchment behavior are caused by several factors *e.g.* the influence of anthropogenic factors combined with the impacts of climate variability on hydrological processes, the methodology used in this study cannot be able to separate the various layers of the synergistic influences in a systematic way for identifying the drivers of flow changes.

6.5 Conclusions and relevance of the findings

6.5.1 Answers to research questions

Can the variability in daily overland flow be adequately explained by the variation in daily rainfall? Are there trends in the model residuals? Was there a change in the response of catchment (as a system) to hydro-climatological inputs?

There was close agreement between the temporal variation of observed and model-based extreme events. The H_0 (no trends in residuals) was rejected at α_s of 5% for all the models used for flow change investigation. The possible change in catchment behavior in its response to rainfall input was minimal. In other words, the flow-rainfall relationship over the data period 1950-2002 was not affected to a large extent by the other factors (apart from rainfall) which might influence the functioning of hydrology.

6.5.2 Relevance of the findings

Like for rainfall, extreme peak flow events also occur in a clustered way in time. The frequency of flow quantiles varies from the periods of "below-normal" and "above-normal" hydrological conditions. This information is relevant for water management, related water supply activities, agriculture, *etc*, as explained before. The variation in flow quantiles over the OHs and OLs can be related to the ERHI over corresponding periods. The findings from this study show that insights on the extreme flow variability can be obtained from the possible drivers of ERHI or dry conditions even if other datasets such as landcover, landuse, soil type, *etc* are missing. This is relevant for the study area given its data scarcity problem.

The remaining amount of flow variability that could not be explained by rainfall variation may be thought of in terms of anthropogenic factors. Such an information can be useful to put in place plans to deal with the influence of the anthropogenic factors on hydrology *e.g.* through local community sensitization, water catchment restoration, and regulation of the management strategies for river banks, mountainous areas, *etc*. However, for more actionable results than from this study, analysis to show extents of human factors across the Nile basin can be made in future research using high quality satellite images of land cover with high spatial and temporal resolutions (if such data become available in the future).

Chapter Seven

Uncertainty in sampling hydro-climatic extremes

This chapter is based on the papers:

Onyutha C, Willems P (2015) Uncertainty in calibrating generalised Pareto distribution to rainfall extremes in Lake Victoria basin. Hydrology Research, 46(3):356–376

Onyutha C, Willems P (2013) Uncertainties in flow-duration-frequency relationships of high and low flow extremes in Lake Victoria basin. Water, 5: 1561-1579

7.1 Introduction

For proper management of water resources under changing climatic conditions and anthropogenic influences, there is an inevitable need for an accurate descriptive study of hydro-meteorological extremes and their recurrence rates or return period T based on long-term time series of observations of rainfall intensities, discharges or water levels, *etc.* Frequency analyses for both high and low flow extremes are relevant for water resources applications related to floods and droughts. Examples of applications include reservoir operations, irrigation, hydropower scheduling, industrial planning, flow control for ecological purposes *e.g.*, compensation flow and dilution flow for improving the quality of water for treatment plants or power generating plants.

It is vital to note that data limitation of the historical time series is a common setback to hydro-climatic extreme value analysis. This creates uncertainty in the estimation of quantiles for planning, design, operation and management of water resources. The quantification and reduction of predictive uncertainty is a crucial issue in statistical modeling of a hydro-climatic variable that allows judgment of the degree of confidence in estimated results. It needs to be taken into account in decision making related to risk-based water engineering and management. Some of the ways for dealing with uncertainty in quantile estimates include regionalization and the at-site approach (Yang *et al.*, 2010).

Regionalization attempts in the Equatorial region of the Nile basin were made by Kizza *et al.* (2013) on the parameters of a conceptual water balance model to estimate catchment inflows, Nyeko-Ogiramoi *et al.* (2012) on an elusive search for regional flood frequency estimates, and Onyutha and Willems (2015a) on empirical-statistical means of estimating flood quantiles from physiographic characteristics. Uncertainty analysis of flood quantile estimates at a regional level was also carried out by Mkhandi *et al.* (1996) for only Tanzania that shares the Southern part of the LVB. However, the accuracy of quantiles obtained for ungauged watersheds based on a regional analysis depends on the density of the stations and, above all, the quality of the data used. This may limit the validity of the regionalization procedures in real life applications. This means that it is important to first deal with the at-site uncertainties, as done in this study, before the data can be used for regionalization procedures for ungauged catchments. Such at-site uncertainty studies in the LVB were carried on rainfall-runoff modeling by Kizza *et al.* (2011). However in this study, uncertainty on sampling peak flow and rainfall extremes was considered.

Several approaches exist to assess errors in statistical models, in which the Monte Carlo simulations is commonly used. Other approaches are the Jackknife method (Tukey, 1958) and the typical-value principle (Hartigan, 1969), which both construct CIs by employing subsample values of a general statistic as the building block. Bootstrapping was introduced by Efron (Efron, 1979) to further strengthen the jackknife method of estimating bias or standard error.

Hydro-meteorological extreme value analysis is commonly based on Partial-Duration-Series (PDS) or Peak-Over-Threshold (POT) extremes extracted from the available series of historical flow and hydro-climatic observations. Following the extreme value theory by Pickands (1975), Generalized Paterno Distribution (GPD) can be calibrated to these extremes (see *e.g.* Willems *et al.*, 2007; Li *et al.*, 2013; Madsen *et al.*, 1997a; Madsen *et al.*, 1997b; Skaugen 2002, Kjeldsen *et al.*, 2000). In frequency analysis, the reliability of the theoretical quantiles estimated from the extreme value distribution (EVD) obviously depends on the accuracy of the GPD

calibration. This is, however, challenged by the proper choice of the type of GPD (normal, heavy or light tail) and by the selection of the best parameter estimation method. Different parameter estimation methods for GPD are known including, among others, the Method of Moments (MOM), L-moments (Lmom), maximum likelihood (ML), and the Weighted Linear Regression in quantile plots (WLR).

This study was divided into two parts. In the first part, the objectives were to: 1) determine the most suitable GPD class, 2) find the best performing parameter estimation method, and 3) assess, based on (1) and (2), the uncertainties in estimating extreme rainfall intensity quantiles for the LVB. Sampling uncertainty of quantile estimates were addressed through Monte Carlo simulations. The second part of the study focused on 1) modeling of the flow-duration-frequency relationship, and 2) quantification of the uncertainties on daily flow quantiles.

Because daily rainfall and flow series were only available for the LVB, this study focused on the hydro-climatic extremes in the Equatorial region but not the entire Nile basin.

7.2 Uncertainty in rainfall intensity quantiles

7.2.1 Methodology

1) Selection of independent extremes

For frequency analysis, the data are required to be independent and identically distributed. This is normally achievable through extraction from full series, either Annual Maxima Series (AMS) or PDS and/or POT extreme events. In AMS, the time slice is chosen to be one hydrological year. Although extreme events obtained from the AMS model show strong independence, the limitation of their number tends to lower the accuracy of extreme value analysis. The PDS or POT model ensures that the most severe events are selected, independent of the time of occurrence. Hence, it intuitively provides a more consistent definition of the extreme value region than the AMS approach. Comparison of the relative performance of

the AMS and PDS/POT approaches for rainfall series can be found in Madsen *et al.* (1997a), and Madsen *et al.* (1997b). According to Madsen *et al.* (1997a), PDS is more efficient than the AMS in estimation of flood quantiles. Madsen *et al.* (1997b) concluded that in general, one should use PDS with exponentially distributed magnitudes for close to zero shape parameter. It was shown that PDS yields more precise quantile estimates than AMS (Yevjevich and Taesombut 1979, Martins and Stedinger 2000). An overview of the recent advances in PDS/POT modeling of extreme hydrological events can be found in the paper by Rosbjerg and Madsen (2004). These authors summarize important extensions of the PDS/POT method since the mid-1990s. They showed PDS model to be competitive with that of the AMS model and highly efficient for estimation of quantiles. Therefore, given the above support for PDS, it was adopted for this study.

In this study, independent rainfall intensity extremes were selected using independence criteria based on threshold values for the time between two successive independent rainfall intensity peaks, the ratio of the minimum rainfall intensity between the two peaks over the peak value, and the peak height (Willems, 2009). The independence criteria are merged from a number of concepts. Firstly, subsequent rainfall intensity peaks can be considered largely independent if the inter-event time exceeds the recession time and if the lowest inter-event intensity drops below a specific low intensity level. Secondly, in line with statistical extreme value analysis, the POTs are selected as the highest events in periods where the hydro-climatic variable up crosses a selected threshold. The threshold component of the method presented by Willems (2009) to extract independent extreme events is based on the paper by Lang *et al.* (1999) which reviews tests and methods useful for modeling the process of POT values, the choice of the threshold level, the verification of the independence of the values and the stationarity of the process.

2) GPD and calculation of the quantile for a given return period

Whereas the extreme events from AMS follow the Generalized Extreme Value (GEV) distribution of Jenkinson (1955), it has been shown in some asymptotic sense that for independent rainfall intensity extremes as used

in this study, the conditional distribution of these extremes follows the GPD of Pickands (1975).

Considering x_t , α and γ as threshold (analogous to location, ξ), scale and shape parameters respectively, the cumulative distribution function $G(x)$ of the GPD is given by:

$$G(x) = 1 - \left\{ 1 + \gamma \frac{(x - x_t)}{\alpha} \right\}^{\frac{-1}{\gamma}}, \text{ for } \gamma \neq 0 \quad (7.1)$$

$$G(x) = 1 - \exp \left\{ - \left(\frac{(x - x_t)}{\alpha} \right) \right\}, \text{ for } \gamma = 0 \quad (7.2)$$

This distribution is valid for values of x above the threshold x_t . Based on the shape parameter γ or ($k = -\gamma$), also called the extreme value index, three GPD classes can be identified including normal ($\gamma = k = 0$), heavy ($\gamma > 0$ or $k < 0$) and light ($\gamma < 0$ or $k > 0$) tails. In the case $\gamma < 0$, the GPD has a final right-end point, whereas in the cases $\gamma = 0$ or $\gamma > 0$ the upper GPD tail goes up to infinite values. The GPD for $\gamma = 0$ equals the exponential distribution.

Considering t as the number of observed rainfall intensity events above x_t , w as the sample size of the POT events, R_T as the rainfall intensity corresponding to a given T or T -year quantile, the relationship between the R_T and T based on calibrated GPD is given by (Rosbjerg, 1985):

$$T [\text{years}] = \left(\frac{N}{f} \right) \frac{1}{1 - G(R_T)} \quad (7.3)$$

where N is the data record length in years, and $f = t$ (for WLR method) and $f = w$ (for MOM, Lmom and ML methods).

Considering i as the rank of the POT events ($i = 1$ for the highest), the empirical exceedance probability or T is given by:

$$T [\text{years}] = \frac{n}{i} \quad (7.4)$$

3) Estimation of GPD parameters

Commonly used parameter estimation methods for calibrating the GPD to extreme hydro-meteorological events include the MOM, Lmom and ML methods. The ML method (Fisher, 1922) holds that the best value of a parameter of a probability distribution should be that value which maximizes the likelihood or joint probability of occurrence of the observed sample. The ML method is mainly used because of its desirable mathematical and optimality properties. However, ML estimates can be heavily biased for small samples. The mathematics involved in its usage is also often non-trivial. The MOM approach (Pearson, 1902a; 1902b) follows the rule that the estimators of the (conventional) population moments should be equal to the sample moments. The main advantage of the MOM approach is its undeniable simplicity when moments are available. However, it lacks the desirable optimality properties like for the ML method. The L-moments are certain linear combinations of Probability Weighted Moments (PWMs) that have simple interpretations as measures of the location, dispersion and shape of the data sample (such as is the case for conventional moments). The main advantage of L-moments is that they suffer less from the effect of sample variability compared with the conventional product moment estimators (Hosking, 1990; Stedinger *et al.*, 1993). The MOM, ML and Lmom estimation of GPD parameters were reviewed by Hosking and Wallis (1987). These authors showed that ML estimators do not exist for shape parameter less than -1. They furthermore proposed PWMs to estimate the GPD parameters to improve the asymptotic variance of those based on the MOM approach.

The comparison of these parameter estimation methods for calibrating GPDs was undertaken in a number of studies for different regions of the world (see *e.g.* Deidda and Puliga, 2009; Öztekin, 2005; Hosking, 1990; Castillo and Hadi, 1997). For continuous samples, Deidda and Puliga (2009) found that the performance of the ML was better than those of other methods. However, they remarked that the determination of efficient estimators for rounded-off samples is still an open problem. Öztekin (2005) found for peak flow of 50 different rivers, most of them in Turkey, that the MOM approach was superior over other parameter estimation

methods. According to Hosking (1990), L-moment estimators when compared with those of ML estimators usually show Lmom method to be reasonably efficient. Castillo and Hadi (1997) remarked that the most serious problem with the MOM and PWM methods is that they can produce nonsensical estimates (estimates inconsistent with the observed data). The above summary shows that the suitability of parameter estimation methods may vary from region to region. It is important to note that apart from MOM, Lmom and ML, there were also other methods introduced for calibrating GPDs, for instance the elemental percentile method (EPM) of Castillo and Hadi (1997), and the Principle of Maximum Entropy (POME) proposed by Singh and Guo (1995). The POME approach was formulated by Jaynes (1957a; 1957b) and Jaynes (1982). Castillo and Hadi (1997) showed that EPM performs well compared to existing methods. By comparing with the existing methods, the parameter estimates yielded by the POME were shown by Singh and Guo (1995) to be either superior or comparable for high skewness. The WLR, which is based on the quantile incremental properties of the GPDs (Beirlant *et al.*, 1996), is another method for calibrating GPDs to extreme events. It has been used in a number of studies including Onyutha and Willems (2015a), Csörgö *et al.* (1985), Kratz and Resnick (1996), Schultze and Steinebach (1996), Willems *et al.* (2007), and Willems (2013). The main advantage of WLR is the selection in a visual way in quantile plots, of the most optimal threshold, the extreme value index and the corresponding distribution class of the EVD.

4) *Normal versus heavy tail*

When the aim of the GPD calibration is to extrapolate to extreme quantile estimates (*e.g.* for water engineering design or planning applications), the shape of the GPD's tail is of primary importance. A major attention in this respect has to go to the sign of the extreme value index γ . According to Cai *et al.* (2013) empirical literature has documented that random variables investigated in meteorology and environmental science exhibit generally values of γ around zero (normal tail GPD). In analysis of hydro-climatic variables, the tendency not to make assumption or fix the distribution class as either normal, heavy-tailed or with finite endpoint is necessitated by

estimators of γ without having prior knowledge on its sign. This presents possibility of systematic over/under-estimation in the tail of the EVD. However, not to lose the said generality so soon in this study, γ was computed using MOM, Lmom, ML and WLR methods and the statistical significance of its difference from zero was investigated. This was done based on the calculation of 95% CI on γ estimate. This evaluation of the sign of parameter γ was twofold; firstly for each of the rainfall series individually, and secondly for all the rainfall series collectively. The idea here was that if γ had a value that was not significantly different from zero across the study area, the exponential distribution as a special case of the GPD with $\gamma = 0$ could be adopted for the statistical modeling of the rainfall extremes.

The MOM, Lmom, ML, and WLR methods provide estimators for the different GPD parameters, including the shape parameter γ as follows:

- a) for MOM approach, as reviewed by Hosking and Wallis (1987):

$$S_k = \frac{2(1+\gamma)(1-2\gamma)^{0.5}}{1-3\gamma} \quad (7.5)$$

where S_k is the sample skewness.

- b) for Lmom method (Hosking 1990):

$$\gamma = \frac{3\tau_3 - 1}{1 + \tau_3} \quad (7.6)$$

where τ_3 is the L -skewness.

For the ML method, no estimator can be obtained for the location ξ because the likelihood function is unbounded with respect to ξ as shown by Singh and Guo (1995). These authors applied the constraint $\xi = x_1$ which is the lowest POT sample value, since for this constraint the likelihood function is maximum. For that constraint, the ML estimator of γ (or $k = -\gamma$) can be computed together with the GPD scale parameter α using (Singh and Guo, 1995):

$$\sum_{i=1}^w \frac{\alpha^{-1} \times (x_i - \xi)}{1 + \alpha^{-1} \times \gamma (x_i - \xi)} = \frac{w}{1 + \gamma} \quad (7.7)$$

$$\sum_{i=1}^w \ln \left\{ 1 + \frac{\gamma (x_i - \xi)}{\alpha} \right\} = w\gamma \quad (7.8)$$

For the WLR method, the γ value can be computed as the inverse slope of the empirical POT values in the Pareto quantile plot when γ is positive (Beirlant *et al.*, 1996; Willems *et al.*, 2007). In this plot the logarithmically transformed empirical POT values are plotted *versus* minus exceedance probability *i.e.* $-\ln(1 - G(x))$ in abscissa, and $\ln(x)$ in ordinate. The slope of the estimate for $\gamma > 0$ can be obtained by least square weighted linear regression using Hill weights (Hill, 1975):

$$\gamma_t = \frac{1}{t-1} \left\{ \sum_{i=1}^{t-1} \ln(x_i) \right\} - \ln(x_t) \quad (7.9)$$

Beirlant *et al.* (1996) developed another approach valid for any sign of γ based on the so-called generalized quantile plot:

$$\gamma_t = \frac{\sum_{i=1}^{t-1} \left\{ -\ln\left(\frac{i}{t}\right) \right\}^{-1} \ln\left(\frac{i}{t}\right) \{ \ln(x_t) - \ln(x_i) \}}{\sum_{i=1}^{t-1} \left\{ -\ln\left(\frac{i}{t}\right) \right\}^{-1} \ln^2\left(\frac{i}{t}\right)} \quad (7.10)$$

To compute the GPD parameters for $\gamma = 0$, hence the exponential distribution parameters α and ξ , Eqs. (7.11)–(7.14) apply:

c) for MOM approach:

$$\alpha = \sigma \quad (7.11)$$

$$\xi = \mu - \alpha \quad (7.12)$$

d) for the Lmom:

$$\alpha = 2\lambda_2 \quad (7.13)$$

$$\xi = \lambda_1 - \alpha \quad (7.14)$$

where λ_1 is the first L-moment, λ_2 the second L-moment, μ the sample mean, and σ the standard deviation.

For the ML method as explained above, the ξ was constrained to x_1 . For a known ξ , the estimate of α using the ML method is identical to that of Lmom or MOM. Therefore, in this study, α of the ML method was determined using Eq. (7.11).

For the WLR method, the estimate of α is obtained by a least squares weighted linear regression in the exponential plot, defined as the plot similar to the Pareto quantile plot but with the $\ln(x)$ taken as x in ordinate. This regression estimate constrained to t observations is given by (Csörgö *et al.*, 1985; Beirlant *et al.*, 1996):

$$\alpha_t = \frac{1}{t-1} \left(\sum_{i=1}^{t-1} x_i \right) - x_t \quad (7.15)$$

Eq. (7.15) is again based on the weighting factors proposed by Hill (1975). The optimal threshold x_t can be defined as the POT value above which the Mean Squared Error (MSE) of the regression is minimal. For a selected t value, the MSE of the weighted linear regression in the exponential quantile plot is:

$$MSE_t = \frac{1}{t-1} \left\{ \sum_{i=1}^{t-1} \left\{ -\ln \left(\frac{i}{t} \right) \right\}^{-1} \left(x_i - x_t - \alpha_t \ln \left(\frac{t}{i} \right) \right)^2 \right\} \quad (7.16)$$

For the exponential distribution of Eq. (7.2), Eq. (7.3) transfers to a linear relationship between R_T and $\log(T)$ as in Eq. (7.17):

$$R_T = \alpha(\log(T) - \log(T_0)) + R_{T_0} \quad (7.17)$$

where R_{T_0} is the rainfall intensity at return period T_0 , and T_0 is equal or higher than the empirical return period n/t of the threshold x_t .

Figure 7.1a shows an example of such calibrated normal tailed GPD as linear regression line in the exponential quantile plot. Example of the calculated slope and MSE for selected t of the same station is shown in Figure 7.1b. Because of high fluctuations in the slope of the quantile plot for high thresholds (see Figure 7.1b), stemming from randomness in the available POT values, the slope estimates for these high thresholds have high statistical uncertainty. Instead for very low thresholds the slope estimates might result in pronounced bias because according to Pickands (1975) the slope converges asymptotically for higher thresholds. The selection of optimal threshold values x_t above which the EVD is calibrated is ensured to be at a point above which the MSE of the linear regression is minimal, *i.e.*, within nearly horizontal sections in the plot of the slope *versus* the number of observations above the threshold. In cases where it is possible for the MSE to reach local minima at different threshold orders within the range where the optimal threshold is situated, the EVD is calibrated for the threshold at each of the local minima for visual aid of selecting the most suitable one. For the example of Figure 7.1, the optimal threshold was determined as the POT value with threshold rank $t = 75$ (the 75th highest rainfall intensity value). A linear tail behavior in the exponential quantile plot can be observed towards the higher rainfall intensity values.

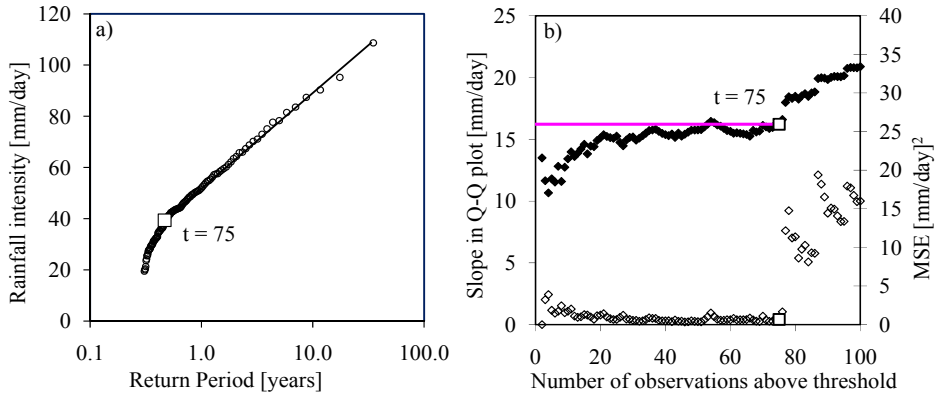


Figure 7.1 a) Observations (○) and calibrated exponential distribution above selected optimal threshold (□) in the exponential quantile plot for daily observed rainfall intensities at Station 2 of Table 2.2, b) Slope (◆) and MSE (◇) for selected t

This procedure for selection of the type of distribution and optimal x_t based on Beirlant *et al.* (1996) can be found in a systematic methodology by Willems *et al.* (2007).

5) Evaluation of calibrated extreme value distribution

i) "Goodness-of-fit" statistics

The "goodness-of-fit" between the empirical and theoretical quantiles was evaluated both statistically and graphically. This was for all the parameter estimation methods done over the same quantile region, *i.e.* for POT events above x_t . Statistically, the Kolmogorov-Smirnov (K-S) (Kolmogorov, 1933; Smirnov, 1936) test (Eq. 7.18) at the significance level α_s of 5% was conducted on the H_0 that the data followed the GPD. The K-S statistic K_{stat} was compared with the critical values (K_{CV}) of the said test. If K_{stat} was found to be greater than K_{CV} , the H_0 was rejected. In this study, the values of K_{stat} for which the H_0 was not rejected were also used to judge on the performance of the parameter estimation methods in the calibration of the EVD to empirical quantiles. The lower the value of K_{stat} for which the H_0 was not rejected at 5% level, the higher the acceptability of the parameter estimation method used to calibrate the EVD.

For all the parameter estimation methods, K_{stat} was computed as:

$$K_{stat} = \text{Max}_{i=1}^t \left\{ G(R_{e,i}) - G(R_{g,i}) \right\} \quad (7.18)$$

where $G(R_{g,i})$ and $G(R_{e,i})$ are the cumulative distributions of theoretical (GPD based) and empirical quantiles respectively.

It is worth noting that the K-S test has well known limitations. K-S only applies to continuous distributions and tends to be more sensitive near the centre of the distribution than in the tails. Ideally the K-S test can be applied to data from two distributions (*i.e.* not the empirical data and the fitted distribution). For the case where distribution is fitted to the data, there is need to adjust, *e.g.* using simulations, the critical values of K_{stat} . However, in this study, the need to adjust the critical values of K_{stat} was neglected because of other "goodness-of-fit" measures considered to complement the K-S test. Eventually, the other statistics included the Standard Error of the Estimate SEE [mm/day], also referred to as the Root Mean Squared Error (RMSE); the *bias* [%], which represents the mean error; and the Probability Plot Correlation Coefficient, $PPCC$ [-]. The $PPCC$ was originally developed for normality testing (Filliben, 1975; Looney and Guldge, 1985) and modified by Vogel (1986). It is a measure of correlation between the theoretical ($R_{g,i}$) and empirical ($R_{e,i}$) quantiles.

$$SEE [mm/day] = \left(\frac{1}{t} \sum_{i=1}^t (R_{e,i} - R_{g,i})^2 \right)^{0.5} \quad (7.19)$$

$$Bias [\%] = \frac{1}{t} \sum_{i=1}^t \left(\frac{R_{e,i} - R_{g,i}}{R_{g,i}} \times 100 \right) \quad (7.20)$$

$$PPCC [-] = \frac{\sum_{i=1}^t (R_{e,i} - \bar{R}_{e,i})(R_{g,i} - \bar{R}_{g,i})}{\left(\sum_{i=1}^t (R_{e,i} - \bar{R}_{e,i})^2 \sum_{i=1}^t (R_{g,i} - \bar{R}_{g,i})^2 \right)^{0.5}} \quad (7.21)$$

where $\bar{R}_{g,i}$ and $\bar{R}_{e,i}$ are the mean values of $R_{g,i}$ and $R_{e,i}$ respectively.

Graphically, the statistics K_{stat} , SEE , $bias$, and $PPCC$ for the EVD calibrated using one parameter estimation method were plotted against those of others to visualize the differences. Performance of the parameter estimation methods was furthermore compared through percentage of times; 1) the H_0 in the K-S test was not rejected, 2) the lowest values of K_{stat} , SEE , and $bias$ were realized, 3) the highest value of $PPCC$ was obtained.

ii) Tail behavior

The calibrated GPDs, moreover, were evaluated graphically in quantile plots, *i.e.* exponential quantile plot for $\gamma = 0$, for checking the shape and accuracy of the slope of the GPD. The key parameter defining the slope in the exponential quantile plot α , obtained using the various parameter estimation methods was graphically compared as well. This was done by plotting values of α obtained by one parameter estimation method against those of others.

iii) Error analysis in the quantile estimation

The distribution of the $bias$ in the extreme quantiles and 95% CI limits on the $bias$ were computed for the various parameter estimation methods assuming that they were random and followed a normal distribution. These were in turn used to determine if the values of $bias$ were statistically significant or not. We also constructed and compared 95% CIs on the rainfall intensity quantiles based on the EVD parameters calculated by the different parameter estimation methods.

The CIs were computed by applying the parametric bootstrapping method. Random samples of equal to number of observed flow extremes above threshold were randomly generated from the EVD using the Monte Carlo method. In total, 1000 synthetic samples were generated. The CIs were computed by the percentile method after ranking the flow quantiles in the generated samples and picking the 25th and 975th quantiles as the upper and lower limits of the 95% CI respectively.

After defining CIs for the entire range of empirical return periods, 10-year rainfall quantile (R_{10}) was considered for comparison of CIs between the various parameter estimation methods. Return period of 10 years was selected because it is commonly applied in planning, design and operation of water resource projects. The difference (*Diff*) was computed between the upper (*UL*) and lower (*LL*) limits of the 95% CI expressed as percentage of R_{10} :

$$Diff [\%] = \frac{UL - LL}{R_{10}} \times 100 \quad (7.22)$$

This *Diff* is indicative of the influence of the parameter estimation uncertainty on the rainfall quantile estimate. It is also worth noting that it can be possible for an empirical quantile to be located outside the CI constructed on the calibrated EVD. A small (large) *Diff* but large (small) *bias* indicates rigidity (flexibility) of the region over which parameters for the calibrated EVD were estimated. The best parameter estimation method would have both its *Diff* and *bias* small.

iv) Uncertainty in empirical extreme quantiles

After selecting the best parameter estimation method using all the datasets (combined observed time series and those from the GCMs), uncertainty in the quantiles estimated only using observed data (excluding the GCMs, in this case) at the different selected stations was assessed. This was done graphically by; comparing empirical *T* versus EVD based *T*, and K_{stat} versus *PPCC*; but also by checking the distribution of the residuals on the modeled *T*, and comparing the CIs on the quantiles. The difference between the GPD parameters at the different stations was also investigated.

7.2.2 Results and discussion

1) Normal versus heavy tail

Figure 7.2 shows collective distribution of γ computed from all the series used in this study. It can be seen that the entire distribution of γ for WLR falls to the right of $\gamma = 0$. This follows from Eq. (7.10), which is valid for $\gamma >$

0 only. In other words, WLR was, here, used to check how close the γ distribution values could be close to those of other calibration methods, and from zero. The differences between zero and the centre of the distribution of γ obtained by WLR, MOM, Lmom and ML methods were 0.072, 0.089, 0.051 and 0.068 respectively. It is also noted from Figure 7.2 that for MOM, Lmom and ML methods, higher fraction of the γ distribution values fell on the positive than negative side. This implied that the possibility of the rainfall series exhibiting light tail for the GPD was far lower than that for heavy or normal tails. This was also expected, given that the light tail case does, not occur that often in hydro-meteorology. Because rainfall does not show upper limits, normal and heavy tail distributions are most often found (see *e.g.* Buishand, 1989; Harremoës and Mikkelsen, 1995; Willems *et al.*, 2007).

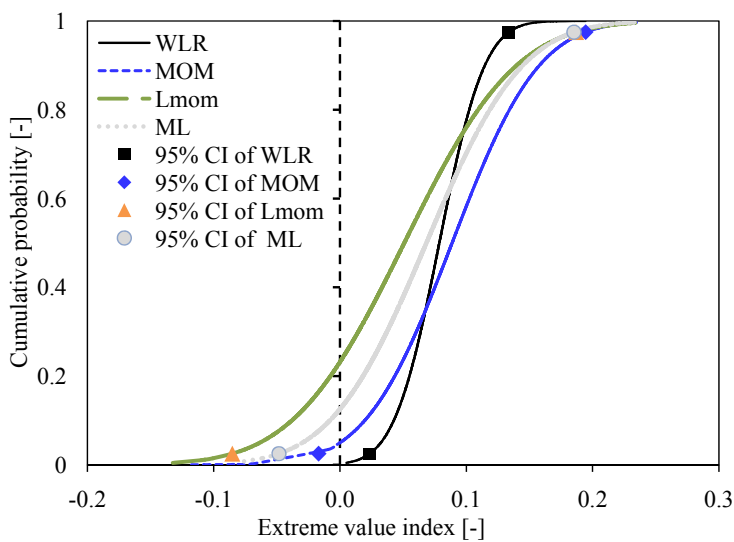


Figure 7.2 Cumulative probability distribution of the extreme value index γ obtained from all the rainfall series (both station-based observed data and those from the GCMs) for the various parameter estimation methods

Figure 7.3 shows the variation in CI for the estimated γ at selected rainfall stations using the different parameter estimation methods. No significant difference is seen between the CIs of the various parameter estimation methods. It is shown from Figure 7.3 that except for WLR, which was

conditioned for $\gamma > 0$, the $\gamma = 0$ case fell within the 95% CI. This meant that the GPD tails did not deviate much from the normal case ($\gamma = 0$). It implied that the GPD could as well be obtainable by approximating γ to zero as also adopted for this study for the extreme value distribution results discussed next.

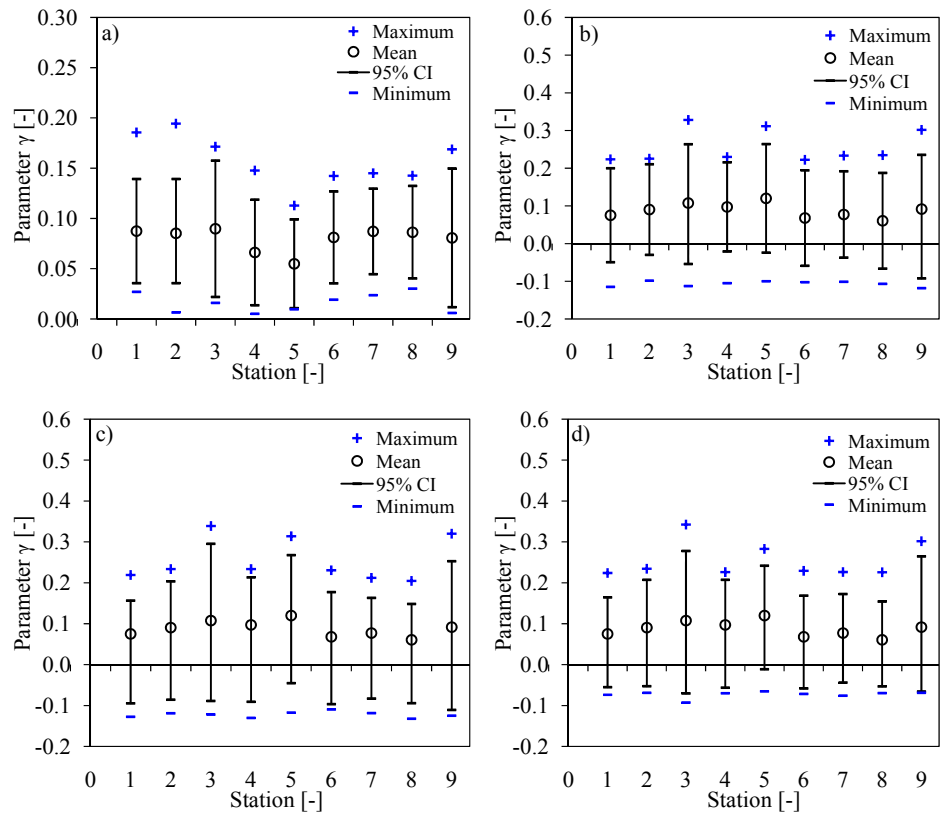


Figure 7.3 Maximum, mean, minimum and 95% CI of the γ distribution obtained from all the rainfall series (both station-based observed data and those from the GCMs) for different stations calculated using: a) WLR, b) MOM, c) Lmom, and d) ML; the label of each station is consistent with that in Table 2.2

2) Evaluation of calibrated extreme value distributions

i) "Goodness-of-fit" statistics

Generally, the EVD parameters estimated by the MOM, ML and Lmom methods were found to be comparable. However, the *bias* on the WLR method was found smaller than those of other parameter estimation approaches. Question was why the *bias* of WLR would be smaller than that of other methods. The best approach to such influences would be to analyze the theoretical *versus* empirical distributions that control the different regions of the plot. Investigation of the tail behavior in the quantile plots (*e.g.* Figure 7.4) shows that the MOM, ML and Lmom methods, although they lead to parameter estimates that give rise to good overall distribution match, may show biased tail behavior (systematically increasing bias towards the higher extremes).

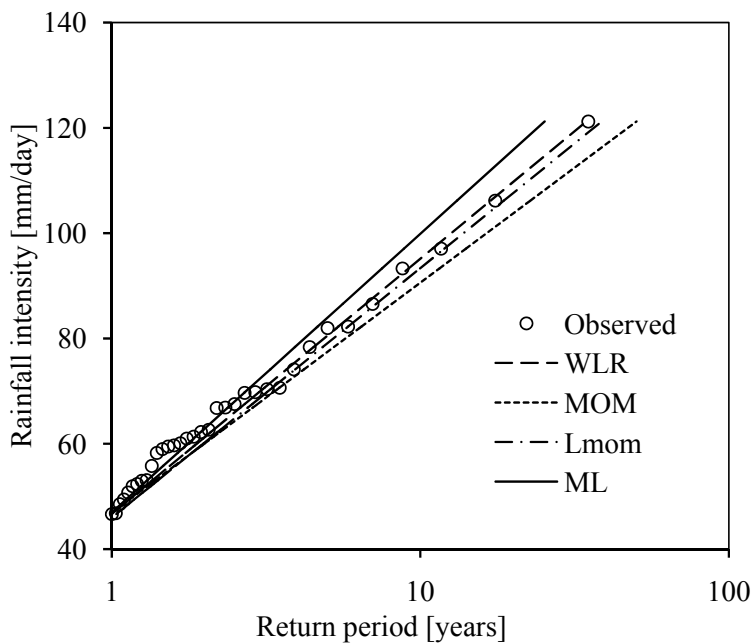


Figure 7.4 Systematic deviations of calibrated exponential distribution from empirical quantiles of Station 4 in Table 2.2

Figure 7.5 shows the probability density of the *SEE* and *bias* for the different parameter estimation methods. Figure 7.5a shows, as already highlighted from Figure 7.4 that the values of *SEE* obtained by WLR were smaller than those of other methods. Figure 7.5b shows the distribution of *bias* in a pattern similar to that of *SEE* in Figure 7.5a. It can be noted that the zero value fell within these CIs for WLR, MOM and Lmom methods; thus their biases were not significant. However, the *bias* for ML was significant. Despite the acceptability of the *bias* in the EVD calibrated using the MOM and Lmom, their uncertainties were higher than those of ML and WLR. These uncertainties were reflected in the difference between the upper and lower limits of the CIs on *bias* as evidenced from Figure 7.5b.

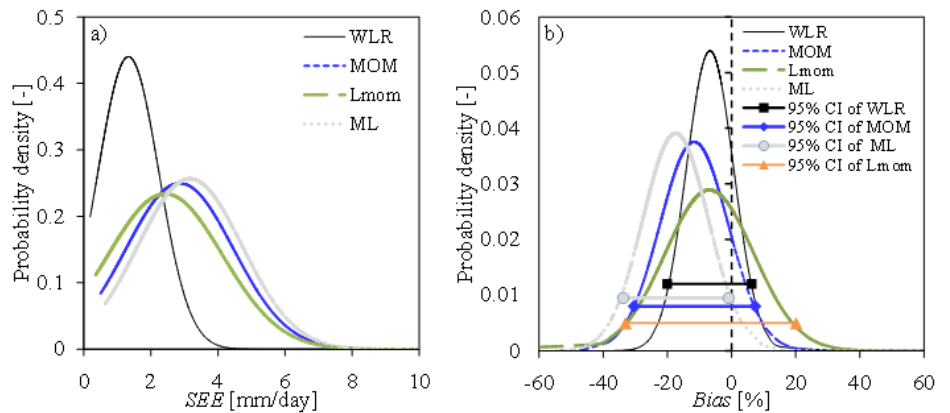


Figure 7.5 Probability distribution of the *SEE* in a) and *bias* in b) on estimated extreme daily rainfall quantiles (from combined series of both station-based observed data and those from the GCMs) for the various parameter estimation methods

ii) Tail behavior

Figure 7.6 shows the differences in the values of parameter α obtained using the various parameter estimation methods. It is shown that the values of parameter α obtained by MOM, Lmom and ML were closer to each other than with those of WLR (Figure 7.6a-c). This can be explained by the difference in the region of POT events for which the parameter α was calculated. The WLR method constrains the computation of parameter α to

POT events above x_t , whereas the computation of parameter α by MOM, Lmom and ML is unconstrained. With increase in parameter α , some systematic deviations (higher α) between the values estimated by Lmom and those by MOM or ML are seen in Figure 7.6d and e. This is due to difference in sensitivity of the mentioned methods to variation or dispersion from the mean of the independent POT events. For the MOM, Lmom, ML methods *versus* WLR, no systematic differences in α estimates are noted.

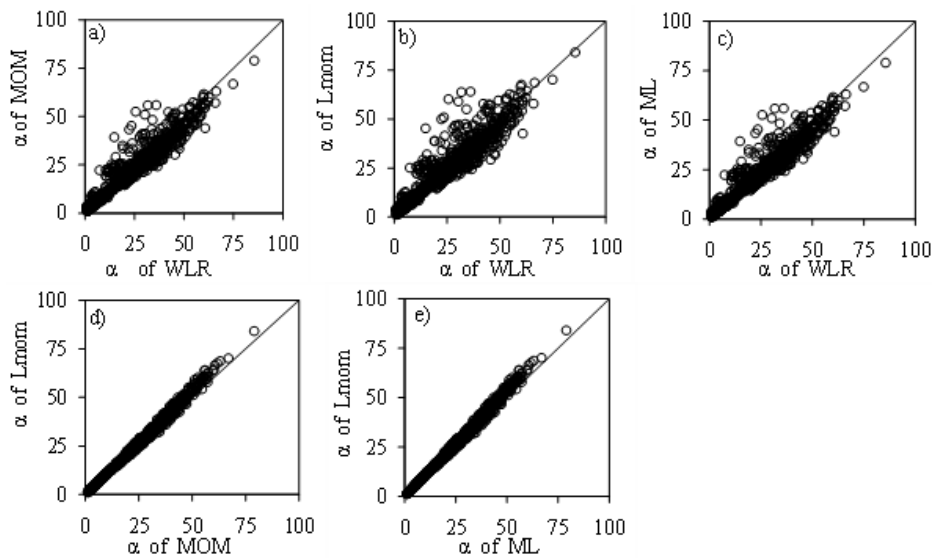


Figure 7.6 Comparison of parameter α obtained by a) WLR and MOM, b) WLR and Lmom, c) WLR and ML, d) MOM and Lmom, and e) ML and Lmom using combined series of both station-based data and those from the GCMs

iii) Summary on evaluation of the calibrated extreme value distributions

Figure 7.7 shows a summary of the performance of the various EVD parameter calibration methods. It is shown that there was no significant difference between the methods in the number of times the H_0 of the K-S test on the calibrated EVD was not rejected. The K-S H_0 was rejected at 5% level in 19.01%, 24.39%, 26.91% and 24.37% of the cases for WLR, MOM, Lmom and ML methods, respectively. The percentages of the cases in which K_{stat} (PPCC) was the lowest (highest) for WLR, MOM, Lmom and ML were

50.70% (58.10%), 17.40% (1.99%), 13.98% (37.92%) and 17.92%(1.99%), respectively. With respect to *Diff*, the Lmom approach presented the smallest value followed by WLR, MOM and finally ML. Based on these results, it was suggested that the most to the least robust method for GPD parameter estimation to rainfall extremes in the LVB with respect to *SEE*, *K_{stat}*, *PPCC* and *bias* were WLR, Lmom, MOM and ML. These results were consistent with those of Mkhandi *et al.* (1996) in a study on the Southern part of the LVB that showed that Lmom could perform better than MOM and ML for flood quantile estimation. However, instead of the conventional calibration of EVDs to the entire POT series, this study suggested that estimations of quantiles using Lmom, MOM and ML methods be obtained with focus of the methods on the EVD tail behavior or region over which GPD (Eq. 7.1 or 7.2) remains valid. This can be achieved when the most optimal thresholds as defined earlier are first selected in a quantile plot, as is the case for the WLR method.

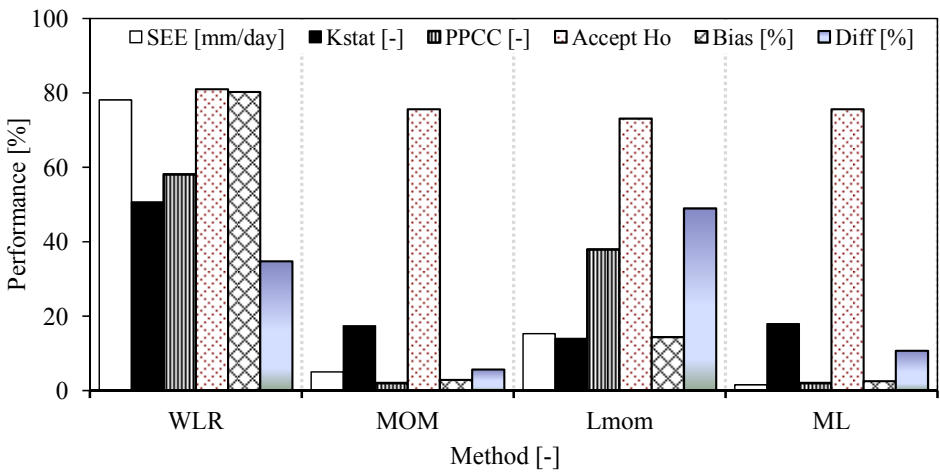


Figure 7.7 Performance of the parameter estimation methods evaluated by the percentage of times the H_0 of the K-S test was not rejected or had the lowest values of *SEE*, *bias*, *K_{stat}* [-], *Diff* or the highest value of *PPCC*

iv) Uncertainty on the empirical extreme quantiles

Figure 7.8 shows, for station-based observed rainfall intensity quantiles, difference in the GPD parameters across the study area. It is shown in

Figure 7.8a that Table 2.2 Stations 3, 4 and 9 (found on the Eastern half) of the LVB were characterized by higher values of the POT threshold x_t than those of 1, 2, and 6-8 (on the Western half). This was because of the higher rainfall extreme intensities in the Eastern half than those in the Western part of the study area. Figure 7.8b shows the scale parameter α of the different stations. No strong difference in the values of α could be seen among Stations 1–2, and 4–8. On the other hand, Stations 3 and 9 are characterized by high values of α . This might could be thought of in terms of the spatial difference in variability.

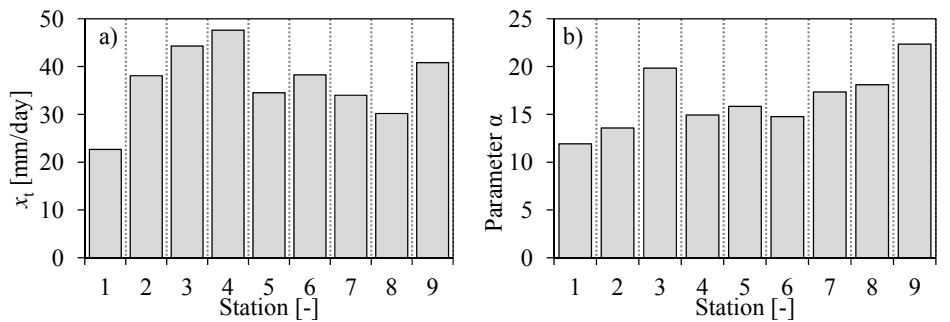


Figure 7.8 GPD parameters, a) x_t , and b) α , estimated by the WLR method using observed rainfall intensity at the different stations; the label of each station is consistent with that of Table 2.2

Figure 7.9 shows, for the WLR method, graphical representations of the uncertainty in modeled T s and T -year events. Figure 7.9a shows comparison between the EVD based and empirical T values. The random deviations were due to data limitations. The systematic deviations of the data points from the bisector showed statistical modeling uncertainty on T and were proportional to the EVD parameter α . Figure 7.9b shows distribution of residuals on modeled T for Station 7 in Table 2.2. It can be seen that the distribution of these residuals followed well the assumed Gaussian distribution; this was found for modeled T at all the selected stations and confirmed by t -tests. Figure 7.9c shows plots the K_{stat} versus $PPCC$. It is shown that $PPCC$ values for all the stations were close to unity. It can also be noted that both statistics led to consistent conclusions: the higher the K_{stat} the lower the $PPCC$. Figure 7.9c confirmed the acceptability of the EVD calibrations to rainfall extremes at all the stations. Figure 7.9d

shows 10-year rainfall intensity quantiles and their estimation uncertainty. The width (difference between upper and lower limits) of the 95% CIs expressed as a percentage of the empirical 10-year rainfall intensity quantile was found to range from 9.25% (at Station 5) up to 59.66% (at Station 3). These uncertainties indicated the need for their consideration in decision making related to risk-based water resources engineering and management. The shifts in the CI between different stations indicated spatial differences in rainfall statistics across the study area. The stations at which wider CIs were obtained (Stations 3 and 9 in Figure 7.9d) present in general higher temporal variability in rainfall intensity. The rainfall extreme intensities in the Nile basin indeed have strong spatial variation (see the regional extreme value analysis by Nyeko-Ogiramoi *et al.* (2012)).

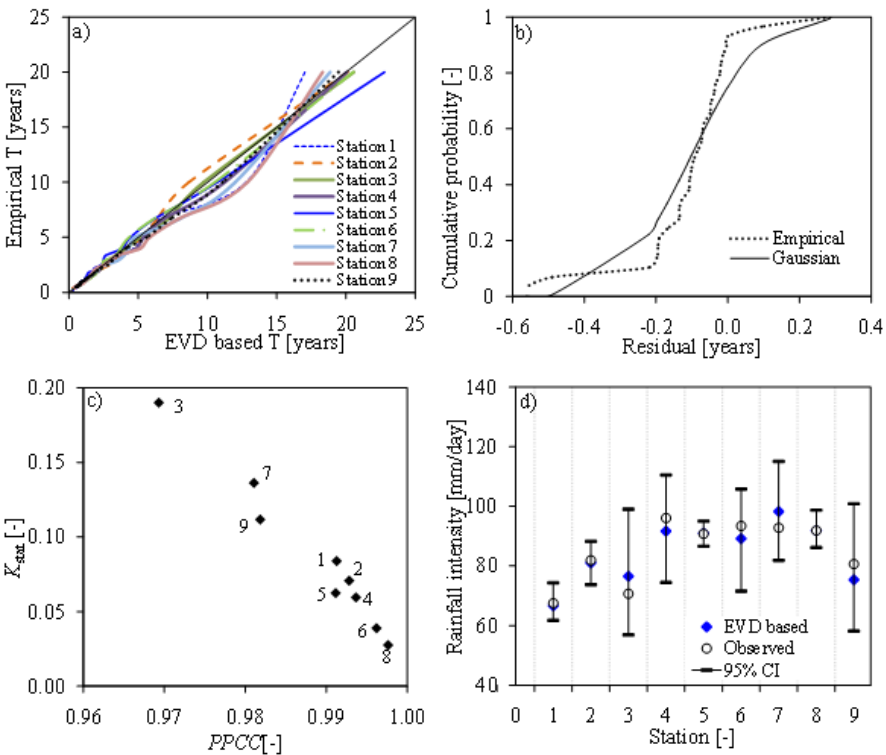


Figure 7.9 a) Empirical T versus EVD based T , b) distribution of residuals on T at Station 3, c) K_{sat} versus PPCC "goodness-of-fit" statistics on calibrated EVD; the labels on the data points refer to the stations of the study, d) 10-year rainfall intensity quantiles and CIs; the label of each station is consistent with that of Table 2.2

7.3 Uncertainty in flow quantiles

7.3.1 Methodology

1) Modeling of Flow Duration Frequency (FDF) relationships

One approach to obtain substantially compressed frequency information on extremes from a hydrological time series is through extreme value analysis for a range of aggregation levels to constitute FDF relationships. Aggregation levels are simply durational intervals over which the hydrological values are averaged. Premised on such durations, the conditional relationships are essentially cumulative functions of the amplitude values in the time series (Chow *et al.*, 1988). FDF relationships are very important in water engineering. According to Nhat (2006), FDF relationships are among the most commonly used tools in water resources engineering, either for planning, designing and operating of water resource projects, or for various engineering projects against floods.

For the FDF modeling, the flow series from Table 2.6 were used. The extreme value analysis and FDF modeling were based on nearly independent high and low flow extremes extracted from the full time series. Independent high flow events were selected using independence criteria based on threshold values as described in Section 7.2.1i. Extraction of low flow events was done by applying the same method but on the inverted flow ($1/F$) series. Prior to the extraction of the extreme values from the full time series for each of the selected stations, moving averaging window was passed through the series. The aggregation levels considered for high flow were 1, 3, 5, 7, 10, 30, 60, 90 days while for low flow analysis 1, 10, 30, 90, 150, 180, 240 and 365 days were taken. This is the range covered by the relevant water engineering or management applications including agriculture, irrigation, hydropower, domestic water supply, pollution control, *etc.* The highest aggregation levels of three months considered for high peak flow events and one year for low flow analysis are based on the differences in time scale of the high/low flow related hydrological processes. Peak flow events are sudden and result in immediate effects due to the excess of water, whereas low flow results

from progressive dry conditions with long-term effects such as shortage in water availability.

Since it is well known in some asymptotic sense that independent POT extreme events conditionally follow the GPD, the identification of the GPD class, estimation of the EVD parameters, and calculation of the quantiles were done based on the WLR technique and quantile plots or Q-Q plots as described in Section 7.2.1 i)-iii). Figure 7.10 shows examples of such calibrated normal tailed GPDs as linear regression lines in exponential Q-Q plots for Station 16 of Table 2.6. The optimal thresholds were determined as the flow values with threshold ranks $t = 100$ and 69 [the 100th and 69th highest flow values for high and low $(1/F)$ flow values respectively] as shown in Figure 7.10a-b. A linear tail behavior in the exponential Q-Q plot can be observed towards the higher F or $(1/F)$ values.

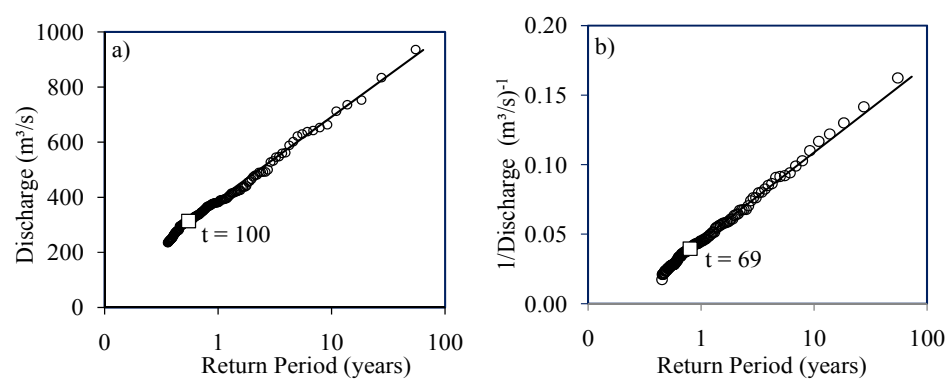


Figure 7.10 a) Daily high flow events; b) Daily low flow events. Symbol (\circ) shows observations in exponential Q-Q plots; (\square) denotes the selected optimal threshold; the regression lines are the calibrated EVDs

What followed next after carefully selecting, in a consistent way, the optimal thresholds for the different aggregation levels, was the calibration of the parameters of the EVD and analysis of the relationship between the model parameters and the aggregation levels. The parameter-aggregation level relationships, together with the analytical description of the EVD, finally constituted the FDF relationships that could be used to estimate

high or low flow quantiles as a simultaneous function of different T s and aggregation levels.

2) *Uncertainty Analysis*

The RMSE and *bias* were computed on the fitted EVDs (for the extreme events above the threshold) using Eqs. (7.19) and (7.20) respectively. The CIs on the quantiles were computed by applying the parametric bootstrapping method as described in Section 7.2.1 iii). After defining CIs for the entire range of empirical return periods, the uncertainty was quantified in terms of the difference between the 95% CI (lower, upper) limits and empirical quantiles (EQ) as percentages of EQ on daily flow events.

7.3.2 *Results and discussion*

1) *FDF relationships*

Figure 7.11 shows examples of the FDF relationships obtained after compiling the exponential EVD calibration results for river flow at various aggregation levels for Station 14 of Table 2.6. Up to T s equal to the length of the available time series, empirical quantiles were derived as well. Because the lengths of the available river flow series were all longer than 25 years but less than 100 years, empirical T -year flow quantiles were obtained only for curves up to 25 years in Figure 7.11a-b. For higher T values, due to the randomness involved in the empirical data, the empirical quantiles can be far more inaccurate in comparison with the theoretical quantiles. Differences between the empirical and theoretical quantiles can, for the higher T values, also be explained by the influence of river flooding and the higher observation errors for higher flow events. One of the reasons of the latter increase in observation errors for higher flow events are due to bias in rating curve extrapolation or the difference between the river discharge and the catchment rainfall-runoff discharge. Another reason is the increasing statistical model uncertainty with increasing T .

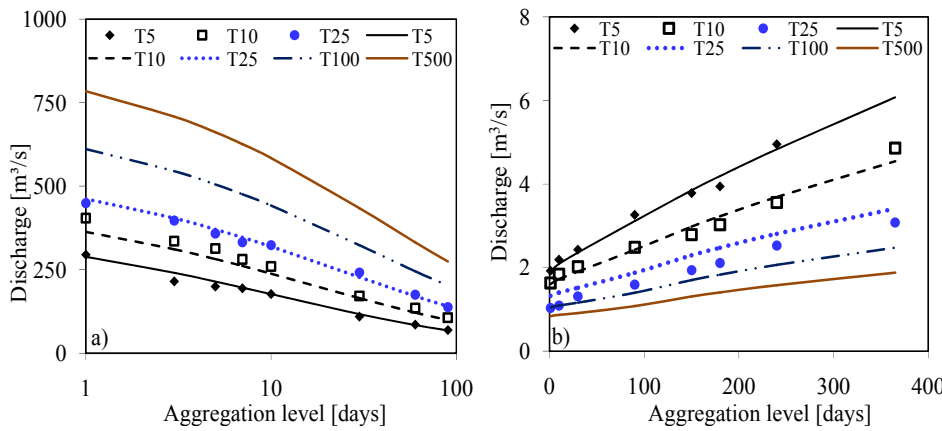


Figure 7.11 FDF relationships of a) high flow, and b) low flow. the horizontal axis of a) is logarithmic; the empirical and theoretical quantiles are shown by markers and lines respectively.

2) *Uncertainty Analysis*

Figure 7.12 shows the bias and RMSE of the deviation between FDF modeled and empirical T averaged over the entire EVD fitted to daily extreme flow events above the threshold. It is noticeable that for both high and low flow events, the average biases for all the selected catchments of the study area were less than 8%. This indicated the acceptability of the exponential distribution to model the extreme high and low flow events of the study area using the WLR technique in Q-Q plots. However, as seen from Figure 7.12, catchments 2, 6 and 7 which are so close to each other in the North Western quadrant of the LVB (see Figure 2.4) exhibited lower bias or RMSE than those of the Western or Eastern portion (see Stations 5, 15, and 24 in Figure 2.4 and Table 2.6). This indicated the reduced variability of the river flow events in this North Western portion compared to other areas of the LVB.

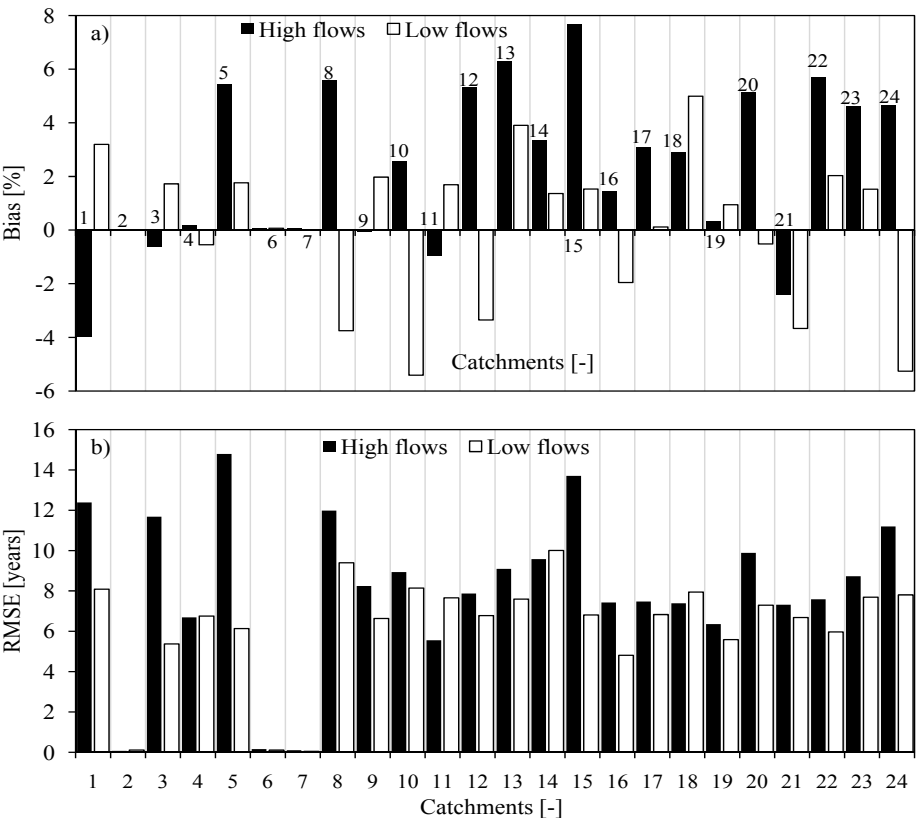


Figure 7.12 a) Bias [%], and b) RMSE on return periods [years]; the labels of the catchments from 1 to 24 are in the order arranged in Table 2.6

Figure 7.13 shows, for illustration using series at Station 16 in Table 2.6, the Monte Carlo simulation results in the form of 95% CI constructed on return periods of high and low hydrological extremes of daily aggregation level. It can be seen that the CIs increase in width with increase in T , which obviously is the result of the high uncertainty associated with extreme value analysis for very high T . Generally, the widths of the CIs were also noted to vary from one catchment to the next thereby reflecting the difference in the degree of temporal variability in river flow events for the different catchments of the LVB.

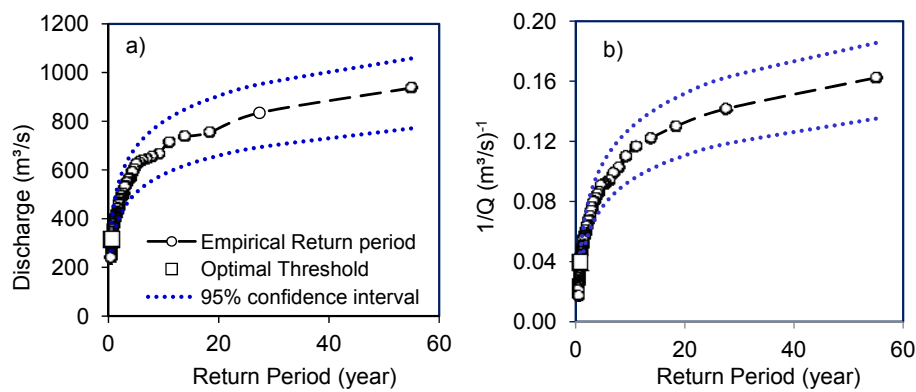


Figure 7.13 Uncertainty on return periods of daily quantiles at Station 16 in Table 2.6 based on daily a) high flow, and b) low flow

Figure 7.14 shows, for illustration, the uncertainty on flow quantiles for return period of 10 years based on the river flow at Nzoia, Rwamba bridge in Kenya (Station 16 in Table 2.6). The average difference between CI (lower, upper) limits and empirical quantiles (EQ) as percentages of EQ on daily flow events of the FDF relationships of high peak flow events were $(-51.9\%, 60.5\%)$, $(-61.0\%, 82.5\%)$ and $(-70.7\%, 116.7\%)$ for T of 5, 10 and 25 years respectively. Correspondingly, for low flow events, CIs of $(-56.1\%, 91.6\%)$, $(-65.5\%, 116.3\%)$ and $(-77.7\%, 151.2\%)$ were obtained. The highest values for the ratio of the 95% CI limit to the EQ (expressed as percentage) were found at Stations 20 and 16 with 155.41% and 179.08% for high and low flow events respectively; and the minimum values were 37.74% and 91.47% at Stations 23 and 11 for high and low flow events respectively. These differences were expected to be due to the variation in the influence of local climate bringing about uneven wet and dry periods across the study area. The differences were noted to become narrower and wider with increase in aggregation levels of high and low flow events, respectively (see Figure 7.14). This explains that the CIs on the FDF based extreme flow quantiles depend on the magnitude of the aggregated river flow events. With increase in the aggregation level, the magnitudes of the river flow events reduced (for high peak flow) and increased (for low flow events) as seen in Figure 7.14a and b, respectively. Importantly, as shown

in Figure 7.14, the CI for any selected T -year quantile on FDF relationship could be estimated.

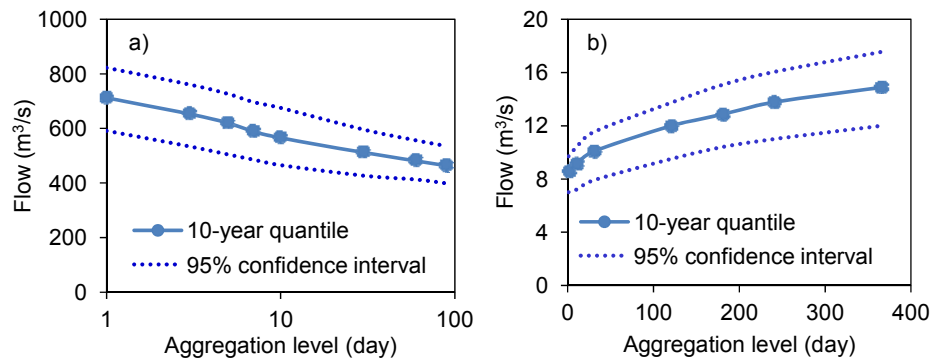


Figure 7.14 Uncertainty on 10-year quantiles of FDF relationship based on a) high flow, and b) low flow at Station 16 in Table 2.6

7.4 Conclusions and relevance of the findings

7.4.1 Answers to research questions

What is the best class of the Extreme Value Distribution (EVD) to model extreme hydro-climatic events of the study area? What is the best method for estimation of the EVD parameters? What is the sampling uncertainty on rainfall and river flow quantile estimates?

The exponential case of the GPD was found suitable for hydro-climatic extremes. For calibrating the EVD, the WLR method was found to be more robust than others. Using daily rainfall intensity (from 1961 to 1990) at 9 locations in the LVB, the difference (*Diff*) between upper and lower limits of the 95% CIs expressed as a percentage of the empirical 10-year quantile was found to range from 9.25% up to 59.66%. For river flow (from 1950 to 2004) at the outlets of 24 catchments in the LVB, *Diff* was, on average, 21.5% and 50.8% for high and low flow respectively.

7.4.2 Relevance of the findings

The findings show high uncertainty in quantile estimates due to data scarcity problems in the study area. There is a need for the decision makers to account for such an uncertainty in planning, design and operation of water resource projects.

Several meteorological and hydrological stations are not continuously operational due to poor maintenance. This brings about the main problem of poor quality of data and/or data limitation across the study area. Eventually, the use of data from such stations leads to huge uncertainty on estimated quantiles. Such an uncertainty lowers the confidence in water resources projects that depend on quantiles. Besides, the uncertainty in estimated quantiles can have serious cost implications for policy makers during project implementation. The findings from this study, therefore, show that the policy makers should invest in hydro-climatic data collection to minimize data limitation and quality problems.

Chapter Eight

Summary and recommendations

8.1 Summary

Rainfall totals, EHRI and dry conditions were found to exhibit oscillatory behavior over multi-decadal time scales. The variability in rainfall of the study area can be explained by the large-scale ocean-atmosphere variability and interactions. Results of variability analyses depend on the selected temporal resolution, data period, and size of the spatial scale. The amount of flow variability that could not be explained by rainfall variation (*e.g.* influence from human factors on hydrology) was found to be small. Uncertainty on hydro-climatic quantiles were not small especially due to data scarcity problems.

8.2 Recommendations

- i) Compared with flow variation analyses based on only four selected catchments in this research, it is recommended that more studies be conducted on many other catchments covering the entire Nile basin. This is to reflect the difference in the influence of the micro-climate (micro-scale features) on the hydrology of the study area.

Furthermore, instead of conceptual models, detailed physically-based hydrological models can be applied to support the simulation-based approach especially if quality of the model inputs improves in the future. Such inputs will include fine-resolution spatial data *e.g.* soil, land-use, *etc* as well as high-resolution and long-term series of observed temperature or evaporation, rainfall, and flow. The use of archives of good quality aerial photos or satellite images of land cover with high spatial and temporal resolutions would also interestingly supplement the flow variability change results from the detailed process-based models. The same applies for inventories (on the dates) of hydraulic structure works, river abstractions and other artificial influences on the river flow.

- ii) It is recommended that alongside the linkage of rainfall variability to large-scale ocean-atmosphere interactions, possible influences from regional features such as water bodies, changes in land cover or use on the spatio-temporal differences in rainfall across the Nile basin should be investigated in a detailed study. To do so, it is envisaged that the use of numerical models (*e.g.* Regional Climate Models) would be important to complement the data-based analysis and determine these physical features. A typical mechanism that should be clearly reproduced is the airflow patterns over the study area under the influence of regional features through a high-resolution representation of the Great Lakes and mountains. Upon a clear representation of such a mechanism, changes in hydrology based on the transition in land-use and/or land cover could be examined. Using scenario land-use changes, possible anthropogenic influences on regional climate could also be experimented.
- iii) The challenge that the contemporary water managers are faced with in their decision making is debatably the impact of climate change on the hydro-climatology. Climate change-related studies for the Nile basin have been confined to catchment-scale. It is therefore recommended that the response of the hydro-climate of the future climatic conditions be investigated considering the entire Nile basin for regional planning of water resources applications in a climate change context. This should be done while explaining possible mechanisms of the future changes in the rainfall over the study area. Some interesting questions to answer are: How well do the GCMs agree, under increasing emission of greenhouse gases, on the current and future representation of wind circulation, extent of the influence from the Indian Ocean on the variability of rainfall over the study area, *etc.*
- iv) Apart from the influence of the selection of spatio-temporal scales, other factors (*e.g.* choice of the methods for analysis, choice of driver, *etc.*) which affect the analysis of rainfall and flow variability should be investigated.

Bibliography

- Abtew W, Melesse AM, Dessalegne T (2009) El Niño Southern Oscillation link to the Blue Nile River basin hydrology. *Hydrological Processes* 23:3653–3660
- AghaKouchak A, Habib E (2010) Application of a conceptual hydrologic Model in teaching hydrologic processes. *International Journal of Engineering Education* 26(4):963–973
- AghaKouchak A, Nakhjiri N, Habib E (2013) An educational model for ensemble streamflow simulation and uncertainty analysis. *Hydrology and Earth System Sciences* 17:445–452
- Allan R, Ansell T (2006) A new globally complete monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004, *Journal of Climate* 19:5816–5842
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration – guidelines for computing crop water requirements – FAO Irrigation and Drainage Paper 56, ISBN 92-5-104219-5. FAO – Food and Agriculture Organization of the United Nations, Rome, Italy.
- Beirlant J, Teugels JL, Vynckier P (1996) *Practical analysis of extreme values*. Leuven University Press, Leuven, Belgium, 137pp.
- Beltrando G, Camberlin P (1993) Interannual variability of rainfall in the Eastern Horn of Africa and indicators of atmospheric circulation. *International Journal of Climatology* 13:533–546
- Berhane F, Zaitchik B, Dezfuli A (2014) Subseasonal analysis of precipitation variability in the Blue Nile River basin. *Journal of Climate* 27:325–344
- Bergström S (1976) *Development and application of a conceptual runoff model for Scandinavian catchments*, SMHI RHO 7, Norrköping, Sweden.
- Beven KJ, Binley AM (1992) The future role of distributed models: model calibration and predictive uncertainty. *Hydrological Processes* 6:279–298

- Bewket W, Sterk G (2005) Dynamics in land cover and its effect on stream flow in the Chemoga watershed, Blue Nile basin, Ethiopia. *Hydrological Processes* 19:445–458
- Blackman RB, Tukey JW (1959) *The Measurement of Power Spectra*. Dover Publications, New York, 190pp.
- Block P, Rajagopalan B (2007) Interannual variability and ensemble forecast of Upper Blue Nile basin Kiremt Season precipitation. *Journal of Hydrometeorology* 8(3):327–343
- Box GEP, Cox DR (1964) An analysis of transformations. *Journal of the Royal Statistical Society Series B* 26:211–243
- Bretherton CS, Smith C, Wallace JM (1992) An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate* 5:541–560
- Buishand TA (1989) Statistics of extremes in climatology. *Statistica Neerlandica* 43:1–30
- Cai J-J, de Haan L, Zhou C (2013) Bias correction in extreme value statistics with index around zero. *Extremes* 16(2):173–201
- Camberlin P (1997) Rainfall anomalies in the source region of the Nile and their connection with the Indian summer monsoon. *Journal of Climate* 10:1380–1392
- Camberlin P (2009) Nile basin climates, in: *The Nile: Origin, Environments, Limnology and Human Use, Monographiae Biologicae*, vol. 89, edited by: Dumont HJ, Springer, Dordrecht, The Netherlands, pp. 307–333.
- Castillo E, Hadi AS (1997) Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association* 92(440):1609–1620
- Cherry S (1996) Singular value decomposition analysis and canonical correlation analysis. *Journal of Climate* 9:2003–2009
- Chow VT, Maidment DR, Mays LW (1988) *Applied Hydrology*. McGraw-Hill, New York, USA, 572pp.

- Csörgö S, Deheuvels P, Mason D (1985) Kernel estimators of the tail index of a distribution. *Annals of Statistics* 13(3):1050–1077
- Davidson AC, Hinkley DV (1997) *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, UK, 582pp.
- Deidda R, Puliga M (2009) Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples. *Physics and Chemistry of the Earth* 34:626–634
- DHI (2007) *MIKE11 – A modeling system for rivers and channels. Reference manual*. DHI water & environment, Hørsholm, Denmark, pp. 278–325.
- Dirks KN, Hay JE, Stow CD, Harris D (1998) High-resolution of rainfall on Norfolk Island, Part II: Interpolation of rainfall data. *Journal of Hydrology* 208:187–193
- Diro G, Grimes DIF, Black E (2011) Teleconnections between Ethiopian summer rainfall and sea surface temperature: Part I – observation and modelling. *Climate Dynamics* 37:121–131
- Efron B (1979) Bootstrap methods: Another look at Jackknife. *Annals of Statistics* 7:1–26
- Elmqvist B (2005) Land use assessment in the drylands of Sudan using historical and recent high-resolution satellite data. *The 31st International Symposium on Remote Sensing of the Environment*, June 2004, St. Petersburg, ISRSE.
- Ezekiel M (1930) *Methods of correlational analysis*. New York, John Wiley and Sons.
- FAO (1997) Irrigation potential in Africa: A basin approach. *FAO Land and Water Bulletin*, FAO, Rome, Italy.
- FAO (2001) FAOCLIM 2: world-wide agroclimatic data. *Environment and Natural Resources, No. 5 (CD-ROM) of working papers series*, FAO, Rome, Italy.
- FAO (2016) South Sudan, regional report. Available online via the link http://www.fao.org/nr/water/aquastat/countries_regions/ssd/ [accessed: 14 July, 2017].

- Filliben JJ (1975) The probability plot correlation coefficient test for normality. *Technometrics* 17:111–117
- Fisher RA (1922) The mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A* 222:309–368
- Fontaine B, Janicot S (1992) Wind-field coherence and its variations over West Africa. *Journal of Climate* 5:512–524
- Fontaine B, Roucou P, Trzaska S (2003) Atmospheric water cycle and moisture fluxes in the West African monsoon: mean annual cycles and relationship using NCEP/NCAR reanalysis. *Geophysical Research Letters* 30(3):3–6
- Fowler HJ, Archer DR (2005) Hydro-climatological variability in the Upper Indus basin and implications for water resources. *The Seventh IAHS Scientific Assembly for the S6 symposium on Regional Hydrological Impacts of Climatic Change-Impact Assessment and Decision Making*, April 2005, Foz do Iguaçu, Brazil, IAHS Publ. 295.
- Gebrehiwot SG, Seibert J, Gärdenäs AI, Mellander PE, Bishop K (2013) Hydrological change detection using modeling: Half a century of runoff from four rivers in the Blue Nile basin. *Water Resources Research* 49:3842–3851
- Gleick PH, Adams DB (2000) *Water: the Potential Consequences of Climate Variability and Change for the Water Resources of the United States*. Pacific Institute for Studies in Development, Environment and Security 654 13th Street Preservation Park Oakland, CA, 162pp.
- Grist JP, Nicholson SE (2001) A study of the dynamic forces influencing rainfall variability in the West Africa Sahel. *Journal of Climate* 14:1337–1359
- Harremoës P, Mikkelsen PS (1995) Properties of extreme point rainfall I: results from a rain gauge system in Denmark. *Atmospheric Research* 37:277–286
- Harrigan S, Murphy C, Hall J, Wilby RL, Sweeney J (2014) Attribution of detected changes in streamflow using multiple working hypotheses. *Hydrology and Earth System Sciences* 18:1935–1952

- Hill BM (1975) A simple and general approach to inference about the tail of a distribution. *Annals of Statistics* 3(5):1163–1174
- Hinkley DV (1977) Jackknifing in unbalanced situations. *Technometrics* 19(3):285–292
- Hoell A, Shukla S, Barlow M, Cannon F, Kelley C, Funk C (2015) The forcing of monthly precipitation variability over Southwest Asia during the Boreal Cold Season. *Journal of Climate* 28(18):7038–7056
- Hosking JRM (1990) Lmoments: Analysis and estimation of distribution using linear combinations of order statistics. *Journal of the Royal Statistical Society B* 51:105–124
- Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29(3):339–349
- Horel JD (1982) On the annual cycle of the tropical Pacific atmosphere and ocean. *Monthly Weather Review* 110:1863–1878
- Horel JD (1984) Complex principal component analysis: Theory and examples. *Journal of Climate and Applied Meteorology* 23:1660–1673
- Hulme M (1992) Rainfall changes in Africa: 1931–1960 to 1961–1990, *International Journal of Climatology* 12:685–699
- Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation. *Science* 269:676–679
- Indeje M, Semazzi HFM, Ogallo LJ (2000) ENSO signals in East African rainfall seasons. *International Journal of Climatology* 20:19–46
- IPCC (2001) *Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson, CA. Cambridge University Press, Cambridge, United Kingdom and New York, USA, 881pp.

- IPCC (2007) *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, Cambridge University Press, Cambridge, United Kingdom and New York, USA, 996pp.
- IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. Cambridge University Press, Cambridge, United Kingdom and New York, USA, 1535pp.
- Jaynes ET (1957a) Information theory and statistical mechanics, I. *Physical Review* 106(4):620–630
- Jaynes ET (1957b) Information theory and statistical mechanics, II. *Physical Review* 108(2):171–190
- Jaynes ET (1982) On the rationale of maximum entropy methods. *Proceedings of the Institute of Electrical and Electronics Engineers* 70:939–952
- Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81:158–171
- Jones PD, Jónsson T, Wheeler D (1997) Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland. *International Journal of Climatology* 17:1433–1450
- Jones PD, Salinger MJ, Mullan AB (1999) Extratropical circulation indices in the Southern Hemisphere based on station data. *International Journal of Climatology* 19:1301–1317
- Jury MR (2010) Ethiopian decadal climate variability. *Theoretical and Applied Climatology* 101:29–40
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M,

- Saha S, White G, Woollen J, *et al.* (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77(3):437–471
- Kendall MG (1975) *Rank correlation methods*. Fourth ed. Charles Griffin, London.
- Kennedy & Donkin Power Ltd in Association with Sir Alexander Gibb & Partners and Kananura Melvin Consulting Engineers (1997) Hydropower Development Plan Part 1 (Final Report) –Volume 8, and Environmental Impact Assessment (Stage 1)–Volume 8, Uganda Electricity Board, Kampala, Uganda.
- Kibiiy J, Kivuma J, Karogo P, Muturi JM, Dulo SO, Roushdy M, Kimaro TA, Akiiki JBM (2010) *Flood and drought forecasting and early warning. Nile Basin Capacity Building Networks (NBCBN)*. Flood Management Research Cluster, Nile Basin Capacity Building Network (NBCBN-SEC) Office, Cairo, Egypt, 68pp.
- Kizza M, Rodhe A, Xu C-Y, Ntale HK, Halldin S (2009) Temporal rainfall variability in the Lake Victoria basin in East Africa during the twentieth century. *Theoretical and Applied Climatology* 98:119–135
- Kizza M, Rodhe A, Xu C-Y, Ntale HK (2011) Modelling catchment inflows into Lake Victoria: uncertainties in rainfall–runoff modelling for the Nzoia River. *Hydrological Sciences Journal* 56(7):1210–1226
- Kizza M, Guerrero J-L, Rodhe A, Xu C-Y, Ntale HK (2013) Modelling catchment inflows into Lake Victoria: regionalisation of the parameters of a conceptual water balance model. *Hydrology Research* 44(5):789–808
- Kjeldsen TR, Lundorf A, Rosbjerg D (2000) Use of a two-component exponential distribution in partial duration modeling of hydrological droughts in Zimbabwean rivers. *Hydrological Sciences Journal* 45(2):285–298
- Kolmogorov AN (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin.

- Korecha D, Barnston AG (2007) Predictability of June–September rainfall in Ethiopia. *Monthly Weather Review* 135:628–650
- Kratz M, Resnick SI (1996) The qq-estimator and heavy tails. *communications in statistics – stochastic models* 12(4):699–724
- Kundzewicz ZW, Robson AJ (2004) Change detection in hydrological records – a review of the methodology. *Hydrological Sciences Journal* 49:7–19
- Lang M, Ouarda TBMJ, Bobée B (1999) Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225:103–117
- Legesse D, Vallet-Coulomba C, Gasse F (2003) Hydrological response of a catchment to climate and land use changes in Tropical Africa: case study South Central Ethiopia. *Journal of Hydrology* 275:67–85
- Legesse D, Vallet-Coulomb C, Gasse F (2004) Analysis of the hydrological response of a tropical terminal lake, Lake Abiyata (Main Ethiopian Rift Valley) to changes in climate and human activities. *Hydrological Processes* 18:487–504
- Li Z, Brissette F, Chen J (2013) Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modeling in Nordic watersheds. *Hydrological Processes* 27 (25):3718–3729
- Looney SW, Gullette TRJ (1985) Use the correlation coefficient with normal probability plots. *The American Statistician* 39(1):75–79
- Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences* 34:1044–1055
- Lyon B, DeWitt DA (2012) A recent and abrupt decline in the East Africa long rains. *Geophysical Research Letters* 39:L02702, doi:10.1029/2011GL050337
- Madsen H (2000) Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. *Journal of Hydrology* 235:276–288
- Madsen H, Rasmussen P, Rosbjerg D (1997a) Comparison of annual maximum series and partial duration series methods for modeling

- extreme hydrologic events, 1. At-site modeling. *Water Resources Research* 33:747–757
- Madsen H, Pearson P, Rosbjerg D (1997b) Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 2. Regional modeling. *Water Resources Research* 33:759–769
- Mango LM, Melesse AM, McClain ME, Gann D, Setegn SG (2011) Land use and climate change impacts on the hydrology of the upper Mara River basin, Kenya: results of a modeling study to support better resource management. *Hydrology and Earth System Sciences* 15:2245–2258
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13(3):245–259
- Martins ES, Stedinger JR (2001) Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research* 37(10):2551–2557
- Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society* 78:1069–1079
- Mbungu W, Ntegeka V, Kahimba FC, Taye M, Willems P (2012) Temporal and spatial variations in hydro-climatic extremes in the Lake Victoria basin. *Physics and Chemistry of the Earth* 50–52:24–33
- McHugh MJ, Rogers JC (2001) North Atlantic Oscillation influence on precipitation variability around the Southeast African Convergence Zone. *Journal of Climate* 14:3631–3642
- Melesse AM, Bekele S, McCornick P (2011) Introduction: hydrology of the Nile in the face of climate and land-use dynamics, in: Nile River basin: hydrology, climate and water use, edited by: Melesse AM, Springer, Dordrecht, The Netherlands, pp. vii–xvii.
- Menne JM, Durre I, Korzeniewski B, McNeill S, Thomas K, Yin X, Anthony S, Ray R, Vose RS, Gleason BE, *et al.* (2012a) Global Historical Climatology Network - Daily (GHCN-Daily), Version 3 (NOAA National Climatic Data Center).

- Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG (2012b) An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* 29:897–910
- Merz B, Vorogushyn S, Uhlemann S, Delgado J, Hundecha Y (2012) HESS Opinions "More efforts and scientific rigour are needed to attribute trends in flood time series". *Hydrology and Earth System Sciences* 16:1379–1387
- Mkhandi SH, Kachroo RK, Guo SL (1996) Uncertainty analysis of flood quantile estimates with reference to Tanzania. *Journal of Hydrology* 185:317–333
- Mora DE, Willems P (2012) Decadal oscillations in rainfall and air temperature in the Paute River Basin-Southern Andes of Ecuador. *Theoretical and Applied Climatology* 108:267–282
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. *Journal of Hydrology* 10:282–290
- Nhat LM, Tachikawa Y, Takara K (2006) Establishment of intensity-duration-frequency curves for precipitation in the monsoon area of Vietnam. *Annals of Disaster Prevention Research Institute Kyoto University* 49:93–103
- Nicholson SE (1996) A review of climate dynamics and climate variability in Eastern Africa, in: *The Limnology, Climatology and Paleoclimatology of the East African Lakes*, edited by: Johnson TC, Odada EO, Gordon and Breach, Amsterdam, the Netherlands, pp. 25–56.
- Nicholson SE, Entekhabi D (1986) The quasi-periodic behavior of rainfall variability in Africa and its relationship to Southern Oscillation. *Journal of Climate and Applied Meteorology* 26:561–578
- Nicholson SE, Kim J (1997) The relationship of the El-Niño Southern Oscillation to African rainfall. *International Journal of Climatology* 17:117–135
- Ntegeka V, Willems P (2008) Trends and multidecadal oscillations in rainfall extremes, based on a more than 100 year time series of 10min

- rainfall intensities at Uccle, Belgium. *Water Resources Research* 44:W07402, doi:10.1029/2007WR006471
- Nyeko-Ogiramoi P, Willems P, Mutua FM, Moges SA (2012) An elusive search for regional flood frequency estimates in the River Nile basin. *Hydrology and Earth System Sciences* 16:3149–3163
- Nyeko-Ogiramoi P, Willems P, Ngirane-Katashaya G (2013) Trend and variability in observed hydrometeorological extremes in the Lake Victoria basin. *Journal of Hydrology* 489:56–73
- Ogallo LJ (1988) Relationships between seasonal rainfall in East Africa and the Southern Oscillation. *International Journal of Climatology* 8:31–43
- Ogallo LJ (1989) The spatial and temporal patterns of the eastern Africa seasonal rainfall derived from principal component analysis. *International Journal of Climatology* 9:145–167
- Olang LO, Fürst J (2011) Effects of land cover change on flood peak discharges and runoff volumes: Model estimates for the Nyando River basin, Kenya. *Hydrological Processes* 25:80–89
- Onyutha C (2016a) Identification of sub-trends from hydro-meteorological series. *Stochastic Environmental Research and Risk Assessment* 30:189–205
- Onyutha C (2016b) Statistical analyses of potential evapotranspiration changes over the period 1930-2012 in the Nile River riparian countries. *Agricultural and Forest Meteorology* 226-227:80–95
- Onyutha C (2016c) Statistical uncertainty in hydro-meteorological trend analyses. *Advances in Meteorology* 2016(Article ID 8701617):1-26
- Onyutha C, Willems P (2013) Uncertainties in flow-duration-frequency relationships of high and low flow extremes in Lake Victoria basin. *Water* 5:1561–1579
- Onyutha C, Willems P (2015a) Empirical statistical characterisation and regionalisation of amplitude-duration-frequency curves for extreme peak flows in the Lake Victoria basin. *Hydrological Sciences Journal* 60(6):997-1012

- Onyutha C, Willems P (2015b) Uncertainty in calibrating generalised Pareto distribution to rainfall extremes in Lake Victoria basin. *Hydrology Research* 46(3):356-376
- Onyutha C, Willems P (2015c) Spatial and temporal variability of rainfall in the Nile basin. *Hydrology and Earth System Sciences* 19(5): 2227–2246
- Onyutha C, Tabari H, Rutkowska A, Nyeko-Ogiramoi P, Willems P (2016) Comparison of different statistical downscaling methods for climate change rainfall projections over the Lake Victoria basin considering CMIP3 and CMIP5. *Journal of Hydro-environment Research* 12:31–45
- Onyutha C, Willems P (2017) Influence of spatial and temporal scales on statistical analyses of rainfall variability in the River Nile basin. *Dynamics of Atmospheres and Oceans* 77:26-42
- Osman YZ, Shamseldin AY (2002) Qualitative rainfall prediction models for central and southern Sudan using El Niño-southern oscillation and Indian Ocean sea surface temperature Indices. *International Journal of Climatology* 22:1861–1878
- Öztekin T (2005) Comparison of parameter estimation methods for the three-parameter generalized Pareto distribution. *Turkish Journal of Agriculture and Forestry* 29:419–428
- Pearson K (1902a) On the systematic fitting of curves to the observations and measurements, parts I and II. *Biometrika* 1:265–303
- Pearson K (1902b) On the systematic fitting of curves to the observations and measurements, parts I and II. *Biometrika* 2:1–23
- Pickands J (1975) Statistical inference using extreme order statistics. *Annals of Statistics* 3:119–131
- Philipps J, McIntyre B (2000) ENSO and interannual rainfall variability in Uganda: implications for agricultural management. *International Journal of Climatology* 20:171–182
- Pittock AB (1980) Patterns of climatic variation in Argentina and Chile, I: Precipitation, 1931–60. *Monthly Weather Review* 108:1347–1361

- Pittock AB (1984) On the reality, stability and usefulness of Southern Hemisphere teleconnections. *Australian Meteorological Magazine* 32:75–82
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research* 108:4407, doi:10.1029/2002JD002670
- Rayner NA, Brohan P, Parker DE, Folland CF, Kennedy JJ, Vanicek M, Ansell T, Tett SFB (2006) Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 data set. *Journal of Climate* 19:446–469
- Richman MB (1986) Rotation of principal components. *Journal of Climatology* 6:293–335
- Rientjes THM, Haile AT, Kebede E, Mannaerts CMM, Habib E, Steenhuis TS (2011) Changes in land cover, rainfall and stream flow in Upper Gilgel Abbay catchment, Blue Nile basin – Ethiopia. *Hydrology and Earth System Sciences* 15:1979–1989
- Ropelewski CF, Halpert MS (1987) Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation. *Monthly Weather Review* 115:1606–1626
- Ropelewski CF, Jones PD (1987) An extension of the Tahiti-Darwin Southern Oscillation Index. *Monthly Weather Review* 115:2161–2165
- Rosbjerg D (1985) Estimation in partial duration series with independent and dependent peak values. *Journal of Hydrology* 76:183–195
- Rosbjerg D, Madsen H (2004) Advanced approaches in PDS/POT modeling of extreme hydrological events, in: Hydrology: science and practice for the 21st century, Volume 1, Proceedings of the British Hydrological Society International Conference, edited by: Webb B, Arnell N, Onof C, MacIntyre N, Gurney, R, Kirby C, Imperial College, London, United Kingdom, pp. 217–220.

- Santos JF, Pulido-Calvo I, Portela MM (2010) Spatial and temporal variability of droughts in Portugal. *Water Resources Research* 46:W03503, doi: 10.1029/2009WR008071
- Schreck CJ, Semazzi FHM (2004) Variability of the recent climate of Eastern Africa. *International Journal of Climatology* 24:681–701
- Schultze J, Steinebach J (1996) On least squares estimates of an exponential tail coefficient. *Statistics and Decisions* 14(4):353–372
- Segele ZT, Lamb PJ (2005) Characterization and variability of Kiremt rainy season over Ethiopia. *Meteorology and Atmospheric Physics* 89:153–180
- Seleshi Y, Demarée GR (1995) Rainfall variability in the Ethiopian and Eritrean highlands and its links with the Southern Oscillation Index. *Journal of Biogeography* 22:945–952
- Seleshi Y, Zanke U (2004) Recent changes in rainfall and rainy days in Ethiopia. *International Journal of Climatology* 24:973–983
- Semazzi FHM, Indeje M (1999) Inter-seasonal variability of ENSO rainfall signal over Africa. *Journal of the African Meteorological Society* 4:81–94
- Sheffield J, Goteti G, Wood EF (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate* 19(13):3088–3111
- Shepard D (1968) A two-dimensional interpolation function for irregularly spaced data, in: Proc. of the 23rd National Conference, Harvard College-Cambridge, Massachusetts, 517–523
- Singh VP, Guo H (1995) Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). *Hydrological Sciences Journal* 40(2):165–181
- Skaugen TA (2002) Spatial disaggregation procedure for precipitation. *Hydrological Sciences Journal* 47(6):943–956
- Smirnov NV (1936) Sur la distribution de w_2 (criterium de MR von Mises). *Comptes-Rendus de l'Academic des Sciences* 202:449–452

- Solomatine DP, Shrestha DL (2009) A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research* 45:W00B11, doi:10.1029/2008WR006839
- Staudinger M, Stahl K, Seibert J, Clark MP, Tallaksen LM (2011) Comparison of hydrological model structures based on recession and low flow simulations. *Hydrology and Earth System Sciences* 15:3447–3459
- Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993) Frequency analysis of extreme events, in: Handbook of Hydrology, edited by: Maidment DR, McGraw-Hill, New York, USA, pp. 18.1–18.66.
- Sutcliffe JV, Ducgdale G, Milford JR (1989) The Sudan floods of 1988. *Hydrological Sciences Journal* 34:355–365
- Taye MT, Willems P (2011) Influence of climate variability on representative QDF predictions of the upper Blue Nile basin. *Journal of Hydrology* 411:355–365
- Taye MT, Willems P (2012) Temporal variability of hydroclimatic extremes in the Blue Nile basin. *Water Resources Research* 48:W03513, doi:10.1029/2011WR011466
- Taye MT, Willems P (2013) Identifying sources of temporal variability in hydrological extremes of the upper Blue Nile basin. *Journal of Hydrology* 499:61–70
- TERI (2017) Know climate change. <http://know.climateofconcern.org> [accessed 21:07.2017]
- Thiery W, Davin E, Panitz H-J, Demuzere M, Lhermitte S, Van Lipzig NPM (2015) The impact of the African Great Lakes on the regional climate. *Journal of Climate* 28:4061–4085
- Tierney JE, Smerdon JE, Anchukaitis KJ, Seager R (2013) Multidecadal variability in East African hydroclimate controlled by the Indian Ocean. *Nature* 493:389–392
- Trenberth KE (1997) The Definition of El Niño. *Bulletin of the American Meteorological Society* 78:2771–2777

- Trenberth KE, Hurrell JW (1994) Decadal atmosphere–ocean variations in the Pacific. *Climate Dynamics* 9:303–319
- Tukey J (1958) Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics* 29:614
- UNEP (2006) *Environment for Development: An Ecosystems Assessment of Lake Victoria basin*, UNEP/PASS, edited by: Odada EO, Olago DO, Ochola W, UNEP/Pan African START Secretariat (PASS), Nairobi, Kenya, 200pp.
- van Oldenborgh GJ, te Raa LA, Dijkstra HA, Philip SY (2009) Frequency- or amplitude dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean. *Ocean Science* 5:293–301
- Vogel RM (1986) The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses. *Water Resources Research* 22:587–590
- Wallace JM, Smith C, Bretherton CS (1992) Singular value decomposition of wintertime sea-surface-temperature and 500-mb height anomalies. *Journal of Climate* 5:561–576
- Williams A, Funk C (2011) A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying Eastern Africa. *Climate Dynamics* 37:2417–2435
- Willems P (2009) A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling & Software* 24:311–321
- Willems P (2013) Multidecadal oscillatory behaviour of rainfall extremes in Europe. *Climatic Change* 120:931–944
- Willems P (2014) Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological Extremes – Part 1: Step-wise model-structure identification and calibration approach. *Journal of Hydrology* 510:578–590
- Willems P, Guillou A, Beirlant J (2007) Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying function. *Journal of Hydrology* 338:221–236

- Willems P, Vrac M (2011) Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology* 402:193–205
- Willems P, Mora D, Vansteenkiste T, Taye MT, Van Steenbergen N (2014) Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes – Part 2: Intercomparison of models and calibration approaches. *Journal of Hydrology* 510:591–609
- World Meteorological Organization (1966) *Climatic change*, Tech. Note No. 79, WMO, Geneva, Switzerland, 79pp.
- Yang T, Shao Q, Hao Z-C, Chen X, Zhang Z, Xu C-Y, Sun L (2010) Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River basin, China. *Journal of Hydrology* 380:386–405
- Yeshanew A, Jury MR (2007) North African climate variability. Part 1: Tropical thermocline coupling. *Theoretical and Applied Climatology* 89:25–36
- Yevjevich V, Taesombut V (1979) Information on flood peaks in daily flow series, in: Input for risk analysis in water systems, edited by: McBean E, Hipel K, Unny T, Water Resources Publications, Highland Ranch, Colorado, USA, pp. 171–192.
- Yue S, Wang C (2004) The Mann–Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management* 18:201–218
- Zhang Z, Xu C-Y, El-Tahir ME-H, Cao J, Singh VP (2012) Spatial and temporal variation of precipitation in Sudan and their possible causes during 1948–2005. *Stochastic Environmental Research and Risk Assessment* 26(3):429–441

Information about the author and list of some publications

About the author



Personal information

Name Charles Onyutha
Date of birth 13th Aug 1983
Nationality Ugandan
E-mail conyutha@gmail.com

Education background

| | | |
|-------------------|-----------------|------------------------------|
| Oct 2013-Sep 2017 | KU Leuven | PhD in Civil Eng |
| Sep 2009-Sep 2011 | KU Leuven & VUB | MSc in Water Resour Eng |
| Sep 2003-May 2007 | Kyambogo Univ | BEng in Civil & Building Eng |

Professional experience (excluding non-academic work)

| | | |
|-------------------|--------------------|--------------------------|
| Oct 2013-Oct 2016 | Teaching Assistant | KU Leuven, Belgium |
| Sep 2012-Aug 2013 | Assistant Lecturer | Kampala Int'l. Univ, Ug. |
| Oct 2011-Aug 2012 | Assistant Lecturer | Makerere University, Ug. |
| Aug 2008-Sep 2009 | Teaching Assistant | Makerere University, Ug. |
| Aug 2007-Aug 2009 | Teaching Assistant | Kyambogo University, Ug. |

List of selected peer reviewed publications used as chapters

Onyutha C, Willems P (2017) Space-time variability of extreme rainfall in the River Nile basin. *International Journal of Climatology*, doi:10.1002/joc.5132

Onyutha C, Willems P (2017) Investigation of flow-rainfall co-variation for catchments selected based on the two main sources of River Nile. *Stochastic Environmental Research and Risk Assessment*, doi: 10.1007/s00477-017-1397-9

Onyutha C, Willems P (2017) Influence of spatial and temporal scales on statistical analyses of rainfall variability in the River Nile basin. *Dynamics of Atmospheres and Oceans* 77:26–42

Onyutha C, Willems P (2015) Spatial and temporal variability of rainfall in the Nile basin. *Hydrology and Earth System Sciences* 19(5):2227–2246

Onyutha C, Willems P (2015) Uncertainty in calibrating generalised Pareto distribution to rainfall extremes in Lake Victoria basin. *Hydrology Research* 46(3):356–376

Onyutha C, Willems P (2013) Uncertainties in flow-duration-frequency relationships of high and low flow extremes in Lake Victoria basin. *Water* 5:1561–1579

Principle fine-tuned by a reality check follows a heedful path

